2010 IEEE International Conference on Robotics and Automation
Anchorage Convention District
May 3-8, 2010, Anchorage, Alaska, USA

Automated Rock Recognition with Wavelet Feature Space Projection and Gaussian Process Classification

Hang Zhou, Sildomar T. Monteiro, Peter Hatherly, Fabio Ramos, Eric Nettleton and Florian Oppolzer
Australian Centre for Field Robotics, The University of Sydney, Australia
{h.zhou, s.monteiro, p.hatherly, f.ramos, e.nettleton, f.oppolzer}@acfr.usyd.edu.au

***Abstract*— A crucial component of an autonomous mine is the ability to infer rock types from mechanical measurements of a drill rig. The major difficulty lies in that there is not a clear one to one correspondence between the mechanical measurements and the rock type due to the mechanical noise as well as the variety of the rock geology. This paper proposes a novel wavelet feature space projection approach to robustly classify rock types from drilling data with Gaussian Process classification. Instead of applying Gaussian Process classifier directly to the given measurement pieces, a group of wavelet features are extracted from the neighboring region of a specific data point. Gaussian Process classification is then carried out on the new extracted wavelet features. By putting neighboring data points into consideration rather than dealing with each data point individually, the underlying pattern can be better captured and more robust to noise and data variations. Experimental results on synthetic data as well as varied real world drilling data have shown the effectiveness of our approach.**

I. INTRODUCTION

The motivation of automated rock recognition research conducted in this paper is to extract useful properties such as rock type and strength from the blast hole drilling data, also called measurements while drilling (MWD) data. This is part of a larger project aimed at developing a fully autonomous, remotely operated open pit mine, whose main challenge is to build representations of the in-ground geology to determine the quantity and quality of the minerals of interest. The rock recognition results are highly desired by the mining industry as they provide information that can be used in the optimization of the mine operations as well as mine planing and design [6]. For instance, rock boundary map is important for the blast hole design as well as general strategic planning, and rock strength can be used to adjust the drilling parameters (e.g., rotation speed and penetration rate, etc.) as well as optimizing the explosives loading for blasting.

The MWD data used for rock recognition are the measurements (also called features in classification) collected from sensors equipped on large drill rigs used in mining for blast hole drilling. They are primarily used to control and monitor the drilling process. Figure 1 shows the autonomous blast hole drill rig that collects the MWD data used in this paper. In this work, we classify the MWD data in a proper way so as to relate the drill performance to the physical properties of the rocks being drilled.

The problem is investigated under the framework of Gaussian Process (GP) classification; a state-of-the-art classification approach with great flexibility and well suited for high

Fig. 1. Autonomous blast hole drill rig used for collecting experimental MWD data in this paper.

dimensional data (which is the case for rock recognition with numerous measurements). Other classifiers such as k-nearest neighbor or naive Bayes suffer from the “curse of dimensionality” [2], thus being less effective for high dimensional spaces. Using the prior obtained from the geophysical logging results, a GP classifier is trained to classify the properties of the rocks being drilled (such as rock type and strength) from a group of MWD features as shown in Table I.

TABLE I
MWD FEATURES USED FOR ROCK RECOGNITION IN THIS PAPER.

Index	MWD feature
1	Rotation speed (RS)
2	Pull down rate (PDR)
3	Rotation pressure (RP)
4	Pull down pressure (PDP)
5	Bit air pressure (BAP)

The difficulty in accurately predicting rock types from MWD data lies in two key areas. Firstly, the MWD values are very noisy and are not quite separable between different rock classes. This is especially true when drill rigs operate in a percussion mode where the bit performs both rotational and hammering motions to better fragment the rocks. This introduces a lot of noise in the measurements from the mechanical structure of the machine. In addition, the underlying geology of the rocks varies even within the same rock type. This results in the MWD values of a rock type spanning across a certain range and MWD values of different rock types overlapping with each other. Identifying which characteristics of the signals (measurements) originated from friction, structure vibration, etc, and which are cause by the properties of a specific rock type is the main challenge addressed in the paper.

As an example, Figure 2 illustrates one of the MWD features called "Bit Air Pressure" (BAP). This data was collected from an iron ore mine in Western Australia. Figure 2(a) shows a typical collection of the BAP values from a group of blast holes. Figure 2(b) is a section of the BAP values in Figure 2(a) and the corresponding rock type is shown in Figure 2(c), where the blue dots are the class labels of the rocks and the gray dashed lines sequentially connect the neighboring class labels. It can be seen in Figure 2(c) that there exists a frequent transition of rock classes in a neighboring region of the BAP values which indicates there is not a clear border between different rock types with regard to the MWD feature values. Figure 2 shows that the MWD feature values are not easily distinguishable from different rock types. Therefore, the classification solution should aim to be less sensitive to the uncertainty of the MWD measurements. This uncertainty could possibly be due to the variation of the underlying geological property as well as the noise that would inevitably occur in data acquisition.

In many classification tasks, it is important to choose a proper feature space from a training data to facilitate the class separation. This is the case here, since the original feature space (which is the MWD features) is hard to reliably separate. Thus, we perform some transformations on the data and map it to a new feature space before feeding to the classifier.

The idea of relating drilling measurements to properties of rocks has been studied before [8][7]. Machine learning methods have also been applied to drilling data based rock recognition [9] [10] [11] [6]. None of those approaches have applied an adequate transformation to the drilling data measures before being classified, although in [6], a "smoothing windowing" method was proposed to simply replace the data with the moving average.

Since the MWD data are collected from each individual blast hole, where the geology / rock properties are strongly correlated within each hole, the neighboring MWD data points in the hole should be put into consideration together rather than dealing with each data point individually. Therefore, our proposed solution applies a "sliding window" over the original data and extract some new wavelet decomposition features (low pass coefficients) out of the raw data covered in the "sliding window". The low pass coefficients reflect the signal "signature" of the neighboring region, ignoring the high frequency component. As a result, the underlying pattern can be better captured and robust to noise and data variations.

The key advantage of wavelet transform is that it is capable of conducting multi-resolution analysis by capturing both the frequency and location information, so that the extracted features can better reflect the inherent characteristics of the data and the classification results are less sensitive to the data distribution variation and noise. The low pass wavelet coefficients extracted in our proposed solution can be better modeled by the GP as the GP kernel is essentially smooth and low pass [1].

The remainder of the paper is organized as follows. A brief introduction to GP classification is given in Section II. Details of wavelet feature space projection are described in Section III. In Section IV, experimental results are presented and discussed, followed by a summary of the main conclusions in Section V.

II. GAUSSIAN PROCESS CLASSIFICATION

The model used to recover the class membership function has a significant impact on the recognition accuracy. In this paper, we consider using the state-of-the-art GP model for classification.

A GP is a collection of random variables, any finite number of which have a joint Gaussian distribution [1]. A GP is fully specified by its mean function $\mu(\mathbf{x})$ and kernel function $k(\mathbf{x}, \mathbf{x}')$, i.e., $f \sim \mathrm{GP}(\mu, k)$. With the prior represented by the GP kernel function, GP classification models the posterior directly [1]. The kernel function's hyperparameters can be learned from the training data. The kernel function studied in this paper is the Radial Basis Function (RBF).

Assume we have a data set $\mathcal{D}$ with n observations $\mathcal{D} = \{(\mathbf{x}_i, y_i), i = 1, 2, \cdots, n\}$, where $\mathbf{x}$ is the input vector of dimension m and y is the class label $[-1, 1]$. The input $n \times m$ matrix is denoted as X. Predictions for new inputs $\mathbf{x}'$ are computed from the given training data using the GP model. As described in [1], GP binary classification is performed by first calculating the distribution over the latent function f corresponding to the test case

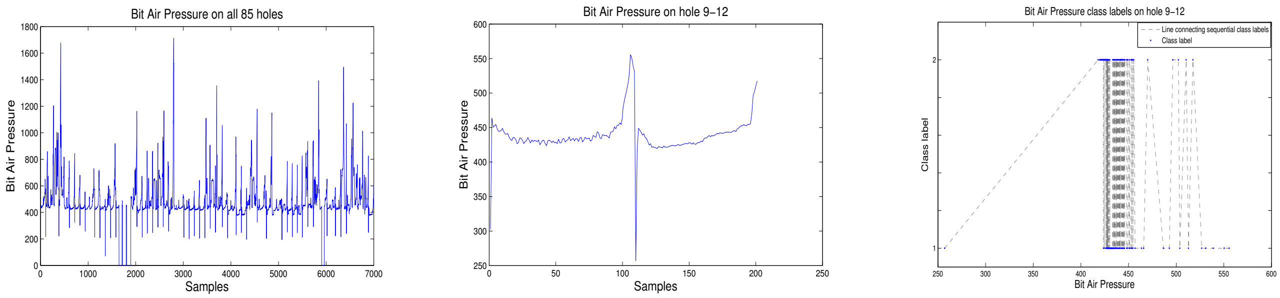
$$p(f'|X, y, \mathbf{x}') = \int p(f'|X, \mathbf{x}', f)p(f|X, y)df \qquad (1)$$

where $p(f|X, y) = p(y|f)p(f|X)/p(y|X)$ is the latent variable posterior, and $p(f'|X, \mathbf{x}', f)$ is the predictive posterior w.r.t. possible latent functions. As the values of this could lie anywhere within the range of $(-\infty, +\infty)$, a second step is necessary to obtain a probabilistic interpretation for the output:

$$\bar{\pi}' = p(y' = +1|X, y, \mathbf{x}') = \int s(f')p(f'|X, y, \mathbf{x}')df' \qquad (2)$$

where s can be any sigmoid function that 'squashes' the prediction output to guarantee a valid probabilistic value within the range of $[0, 1]$.

For the multi-class classification problem with c classes (such as rock recognition), we turn it into a series of $(c(c-1)/2)$ one versus one two-class problems and apply binary classification individually to each of them, followed by majority vote to assign the class labels [4].



(a) Typical "Bit Air Pressure" values from a group of blast holes.

(b) A section of the "Bit Air Pressure" values in (a).

(c) "Bit Air Pressure" labels of the data in (b).

Fig. 2. "Bit Air Pressure" values and labels.

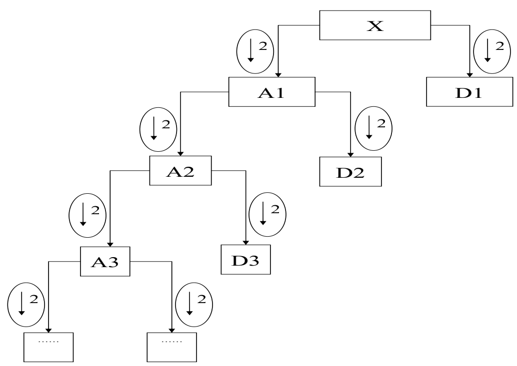


Fig. 3. 1D Wavelet decomposition.

III. WAVELET FEATURE SPACE PROJECTION

A. *Wavelet Transform*

Features are extracted from data through a wavelet analysis before passing to the classifier. Wavelets [12][13] are mathematical functions that decompose data into different frequency components, with the fundamental idea of analyzing according to scale. Wavelet transforms have advantages over traditional Fourier methods in that they are capable of providing time and frequency representations simultaneously while Fourier transforms could only provide frequency. Hence, wavelet transforms are more suitable for analyzing non-stationary data where the signal has time varying frequency, which is normally the case for real world data including the MWD data in our context.

Figure 3 shows a multi-level 1D wavelet decomposition. At each level n, the data is convolved with a low pass filter and a high pass filter. The outputs of both filters are then downsampled to obtain the approximation part An and the detail part Dn respectively.

As indicated in Section I, the main idea of our proposed solution is to capture the signature of the MWD feature values within a neighboring region by extracting a group of low pass wavelet coefficients. In our work, we apply Haar wavelet which is the simplest wavelet transform. The advantage of Haar is that it is fast, yet preserves considerably more details compared with a mean or median filter [3]. On the other hand, the Haar transform also has the limitation of missing the high frequency changes on the high frequency coefficients, due to the two elements wide transform window. Since we only pick up a group of low frequency wavelet coefficients in our approach, Haar is a both efficient and effective choice.

B. *Wavelet Feature Space Projection*

Since the real world MWD drilling data can not be easily distinguished due to the uncertainties caused by possible noise and geological variation (see Section I), our proposed solution deals with a neighboring region of a data point (as shown in Figure 4) rather than individual points of the raw data.

Haar wavelet transform is applied individually to each of the MWD features as shown in Table I. Assume n MWD data points are collected from a blast hole. To extract wavelet features from data on each MWD feature, a window of width $2r+1$ is put on the length n MWD data. We then slide the window from the beginning to the end of the data points. At each position, wavelet decomposition is applied to the data covered in the window. Then, the wavelet coefficients will replace the raw feature data at the center of the window. For a data set with m features, if m' wavelet coefficients are extracted from the "sliding window" on each feature, the transformed data will have a total of $m \times m'$ features sent to the classifier. E.g., for a dataset whose length is 7825 points, if we use 5 of the features of the original

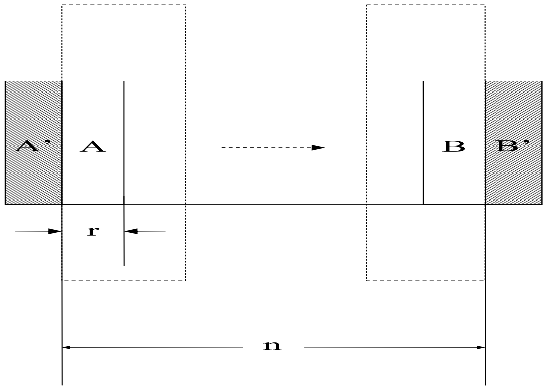
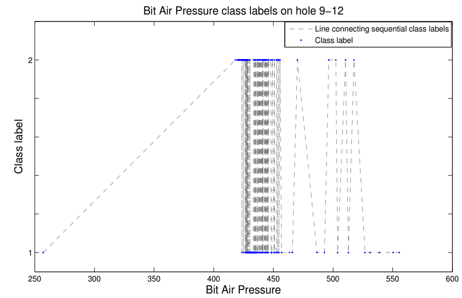
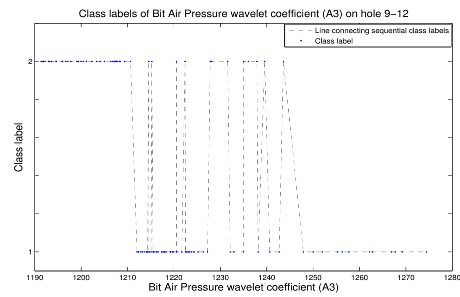


Fig. 4. Sliding window of width $2r + 1$ over length n data for wavelet feature extraction.



(a) Original "Bit Air Pressure" value.



(b) Wavelet coefficients of "Bit Air Pressure".

Fig. 5. Class labels comparison on "Bit Air Pressure".

dataset for classification (which is the case for MWD data in our work), the original data matrix size is 7525×5. If 3 wavelet coefficients are extracted from each feature, the wavelet transformed data matrix size will turn out to be 7525×15.

Fortunately, as indicated in Section I, GP is less sensitive to the "curse of dimensionality". In fact, one of the advantages of GPs is the ability to model sparse data in high dimensional spaces. By processing the data through wavelet transformation with the proposed "sliding window", the inherent multi-resolution characteristics of a neighboring region can be better captured and classification results are much less sensitive to individual signal "spikes".

In our application, we apply 3 to 4 levels of 1D wavelet decomposition to the data on each original input dimension and extract 3 to 5 wavelet coefficients out of each wavelet decomposed feature. In this way, the main underlying characteristics can be captured while still keeping a reasonable number of features for the wavelet transformed data.

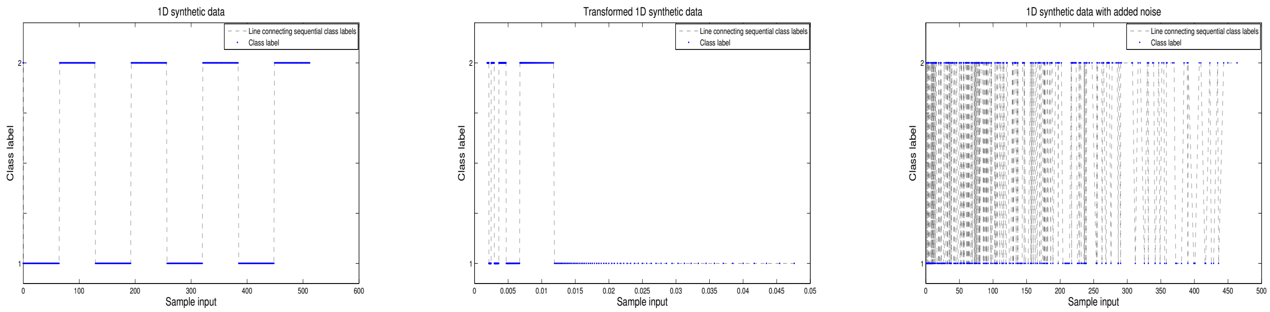
Figure 5 shows a comparison of the BAP (which is one of the original MWD features shown in Table I) value and one of the wavelet coefficients extracted from BAP. It can be seen that wavelet transform nonlinearly "stretches" the data values, making it more separable. Also, from the data distribution point of view, such a "stretch" also makes the data more evenly distributed and hence can better fit to the GP classifier whose kernel is stationary, i.e. shift invariant.

It should be noted that the "sliding window" (with a width of $2r+1$) procedure will chop the beginning and ending part of the data by r each. To compensate, a data segment A with length r can be concatenated to the beginning of the data point as shown in Figure 4 (the shadowed part A'). Likewise for the ending part of the data by connecting segment B next to the ending point (the shadowed part B').

IV. EXPERIMENTS AND RESULTS

We have evaluated our algorithm on both synthetic data and real world MWD data. As described in Section II, the multi-class rock recognition problem is decomposed into a series of one versus one binary classification tasks followed by majority vote to assign the class label [4]. Binary GP classification is implemented using Lawrence's fast GP classification approach [5]. The sliding window half width r is set to be 50 for wavelet feature extraction.

GP classification is tested on the given datasets using k-fold cross validation. k is chosen to be 10 for synthetic data as well as rotary drilling data and 20 for percussion drilling data. The classification results are evaluated by calculating accuracy, precision and recall, which reflect the classification performance from varied aspects. Accuracy is the percentage of all correct predictions (both positive and negative), precision is the ratio of correct labels among the positive predictions and recall is the percentage of the positive labels that has been correctly predicted. In addition, F-measure [14] which is the weighted average of precision and recall is further calculated, so as to integrate precision and recall in one measure to avoid redundancy. Calculation formulae of all the classification performance measures are shown in Table II.



(a) Original data. (b) Transformed data. (c) Data with added noise.

Fig. 6. Synthetic data.

TABLE II
FORMULAE ON ACCURACY, PRECISION, RECALL AND F-MEASURE.

Measure	Formula	
Accuracy (ACC)	TP / (TP+FP)	
Precision (PREC)	TP / (TP+FN)	
Recall (REC)	(TP+TN) / (TP+TN+FP+FN)	
F-measure	2*(PREC*REC) / (PREC+REC)	
	TP: true positive	FP: false positive
	TN: true negative	FN: false negative

A. Synthetic Data

The proposed method was initially tested on synthetic data, where a total of 512 points in 1D and 2 classes are created from rectangular, equal interval pulse as shown in Figure 6(a). The data inputs are the monotonically increased values of 1 to 512 and the output labels are either 1 or -1. To simulate the common situations existing in the MWD data (difficult separation between classes due to possible variations and noise), two transformations are applied to change the data distribution as well as add noise. The first aims to make the data input unevenly distributed by applying Equation (3) (a nonlinear transformation), the second adds noise by multiplying the original data inputs with uniformly distributed random values within the interval $[0, 1]$. The two transformed datasets are shown in Figure 6(b) and Figure 6(c) respectively.

$$x = \frac{1}{x + 20} \tag{3}$$

GP classification with and without wavelet feature space projection are applied on both the original data and the two transformed datasets with the GP classification results shown in Table III which includes accuracies and F-measures for all the three synthetic datasets with and without wavelet projection.

The results show that when classifying directly on the datasets, the accuracy is poor on the original data and further deteriorated on both the nonlinear transformed data and the noise added data. This is mainly because the data do not match with the stationary assumption of a GP model [1] where the neighboring data points are expected to have a smooth transition. By preprocessing the data with wavelet feature space projection (which turns the data towards more stationary by adding correlation among neighboring data points), the GP classification results are improved, and become less variable among different synthetic datasets. This demonstrates that wavelet feature space projection makes the GP classification less sensitive to data distribution variations and noise.

Table III shows that the classification accuracy with the wavelet transformed data is clearly higher (the classification F-measures follow a similar trend). In addition, mean accuracies and F-measures of the classification results across the three synthetic datasets are listed in Table IV. It can be seen from Table IV that on average, classification results of wavelet transformed data consistently outperform the original data on all performance measures. The standard deviations of the wavelet transformed data classification results are significantly lower than those without the wavelet transform, which demonstrates that the wavelet transformation on the data keeps the classification performance stable on varied situations.

B. Real World Drilling Data

Tests were further carried out using various real world MWD data collected from 135 blast holes in an iron ore open pit mine in Western Australia as shown in Figure 7. This includes data from different drilling modes, i.e., rotary and percussion. Each blast hole is around 10 m deep, and the

TABLE III

CLASSIFICATION RESULTS ON SYNTHETIC DATA.

	Data type	GP	GP with wavelet
Accuracy	Original data	0.6328	0.8594
	Transformed data	0.4121	0.8535
	Noise added data	0.3516	0.8414
F-measure	Original data	0.6270	0.8531
	Transformed data	0.4615	0.8472
	Noise added data	0.3616	0.8364

TABLE IV

MEAN PERFORMANCE COMPARISON OF CLASSIFICATION RESULTS ON SYNTHETIC DATA.

Statistics	GP	GP with wavelet
Mean accuracy	0.4655 (0.1480)	0.8514 (0.0092)
Mean F-measure	0.4834 (0.1340)	0.8456 (0.0085)
	* Values in brackets are std.	

MWD feature values in each hole are downsampled at 10 cm intervals. This makes the total number of data points at each hole approximately 100 to 120. As described in Table I in Section I, a total of 5 blast hole MWD features are used for our classification analysis. Mining geologists label and correspond the MWD feature values to several lithological rock types. Two different labeling systems are applied; the first contains 3 classes corresponding to rock types (as shown in Figure 7): shale, iron ore and BIF (banded iron formation); the second contains 5 classes by further subdividing ore and classifying the rocks as shale, low grade medium ore, high grade medium ore, high grade soft ore and BIF. By integrating the drilling modes with the labeling methods, a total of four MWD datasets are used for testing, i.e., rotary - 3 classes, percussion - 3 classes, rotary - 5 classes and percussion - 5 classes.

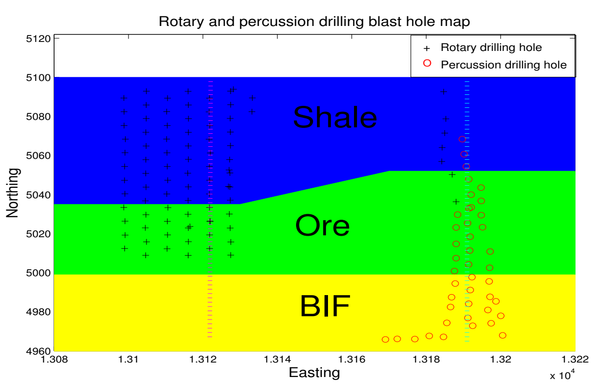


Fig. 7. Map of the blast holes (including the basic geology) from which our testing datasets are collected.

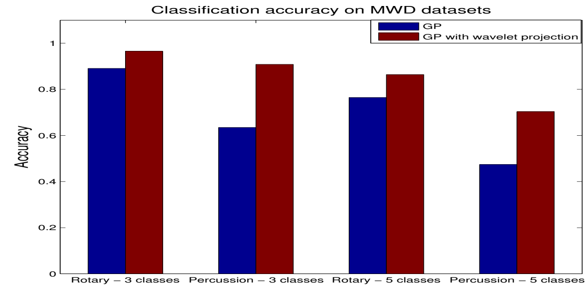


Fig. 8. Classification accuracy on MWD datasets.

TABLE V

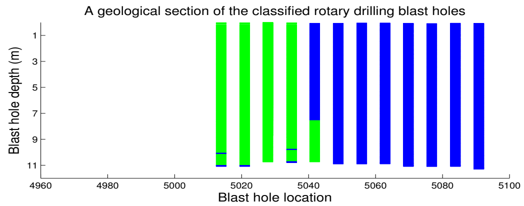
MEAN PERFORMANCE COMPARISON OF CLASSIFICATION RESULTS ON REAL WORLD DRILLING DATA.

Statistics	GP	GP with wavelet
Mean accuracy	0.6909 (0.1788)	0.8601 (0.1127)
Mean F-measure	0.5173 (0.2574)	0.7785 (0.1787)
	* Values in brackets are std.	

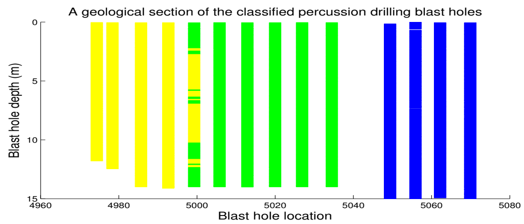
The MWD data classification results are shown in Figure 8. It can be seen that in either rotary or percussion drilling mode and no matter if the labeling is 3 class or 5 class, all the classification evaluation results (accuracy and F-measure) of the wavelet transformed data constantly outperform the classification results on the original data. The obtained results clearly show the stability of the proposed wavelet transformation method on varied real world MWD datasets.

In Table V, mean and standard deviation values of the classification results across all four MWD datasets are presented. The average classification performance of wavelet transformed data across the four MWD datasets well outperform the original data while the relevant standard deviations are all considerably lower than the original data. This again shows that the wavelet transformation makes the classification output more stable on different MWD datasets.

A more intuitive understanding can be provided by considering the results presented in Figure 9. Here, Figure 9(a) is the cross section along the dotted line through rotary blast holes in Figure 7. Figure 9(b) is the cross section cutting through the percussion holes in Figure 7. Relating Figure 9 to Figure 7, it can be seen that the classification results do reflect the underlying geology categories as well as the transitions.



(a) Geological section of the classified rotary (3classes) drilling blast holes.



(b) Geological section of the classified percussion (3 classes) drilling blast holes.

(c) Legend for the figures above.

Fig. 9. Geological sections of the classified (3 classes) blast holes.

C. Discussion

To summarize, our experiments on both the synthetic data as well as the real world MWD data have shown that our proposed wavelet feature projection is effective in providing accurate MWD data based rock recognition. Looking at the synthetic data (in Table III), the improvement of wavelet transform is more prominent on the transformed data as well as the noise added data compared with the original data. In the real world MWD data (in Figure 8), the largest performance leap caused by wavelet projection lies on the percussion data which is usually more noisy (as indicated in Section I). All these show that the wavelet projection approach is robust to data noise and data variation with regard to GP classification. It should also be noted that the improvements mentioned above are achieved by adding very little complexity as the Haar transform used in our approach is both simple and fast.

V. CONCLUSIONS

Building effective representations of the in-ground geology to determine the quantity and quality of the minerals is crucial for large scale mine automation. The work presented here has begun to investigate this issue by proposing a robust rock recognition approach. A new wavelet based data transformation method is presented to project the real world MWD data to a new feature space before applying GP classification. By extracting wavelet features from the neighboring region of a data point, the underlying pattern of the data is more robustly captured, less sensitive to data variation and noise, and better modeled by the GP classifier. The application of our approach is not limited to the MWD data only, but can be generalized to other classification tasks in robotics. Experimental results have clearly shown the advantage of this approach.

VI. ACKNOWLEDGMENTS

This work has been supported by the Rio Tinto Centre for Mine Automation and the ARC Centre of Excellence programme, funded by the Australian Research Council (ARC) and the New South Wales State Government. The authors also acknowledge the support of Annette Pal, James Batchelor and Charles McHugh of Rio Tinto.

REFERENCES

[1] C.E. Rasmussen and C.K.I. Williams, *Gaussian Processes for Machine Learning*, Springer Science+Business Media, LLC; 2006.

[2] C. M. Bishop, *Pattern Recognition and Machine Learning*, The MIT Press; 2006.

[3] A. Jensen and A. LA Cour-Harbo, *Ripples in Mathematics*, Springer; 2001.

[4] J. Friedman, "Another Approach to Polychotomous Classification", *Technical Report*, Stanford University; 1996.

[5] N. Lawrence, M. Seeger and R. Herbrich, "Fast Sparse Gaussian Process Methods: The Informative Vector Machine", *Advances in Neural Information Processing Systems*; 2003.

[6] J. E. J. M. Gonzalez, "Application of Pattern Recognition Techniques to Monitoring-While-Drilling on a Rotary Electric Blasthole Drill at an Open-Pit Coal Mine", *Thesis for Master of Science*, Queen's University, Canada; 2007.

[7] M. J. Scoble, J. Peck, C. Hendricks, "Correlation between Rotary Drill Performance Parameters and Borehole Geophysical Logging", *Mining Science and Technology*, 8, pp. 301-312; 1989.

[8] R. Teale, "The Concept of Specific Energy in Rock Drilling", *International Journal of Rock Mechanics and Mining Sciences*, 2, pp. 57-73; 1965.

[9] K. Itakura, K. Sato, Y. Ichihara, G. Deguchi, H. Matsumoto and H. Eguchi, "Development of a Roof Logging System by Rock Bolt Drilling", *Transactions Institute of Mining and Metallurgy*, Section A 106, pp. 118-123; 1997.

[10] R. L. King, M. A. Hicks and S. P. Signer, "Using Unsupervised Learning for Feature Detection in a Coal Mine Roof ", *Engineering Applications of Artificial Intelligence*, 6(6), pp. 565-573; 1993.

[11] W. K. Utt, "Neural Network Technology for Strata Strength Characterization ", *In Proceedings of International Joint Conference on Neural Networks*, 6, pp. 3806-3809; 1999.

[12] S. Mallat, "Theory for Multiresolution Signal Decomposition: the Wavelet Representation", *IEEE Pattern Analysis and Machine Intelligence*, 11(2), pp. 674-693; 1989.

[13] A. Graps, "An Introduction to Wavelets", *IEEE Computational Science and Engineering*, 2(2), pp. 50-61; 1995.

[14] C. J. van Rijsbergen, *Information Retrieval*, Butterworth; 1979.