Visibility Classification of Pellets in Piles for Sizing Without Overlapped Particle Error

Tobias Andersson   Matthew J. Thurley

Olov Marklund

Luleå University of Technology

EISLAB

Department of Computer Science and Electrical Engineering

Luleå, Sweden,

{tobias.andersson, matthew.thurley, olov.marklund}@ltu.se

Abstract

Size measurement of pellets in industry is usually performed by manual sampling and sieving techniques. Automatic on-line analysis of pellet size based on image analysis techniques would allow non-invasive, frequent and consistent measurement. We make a distinction between entirely visible and partially visible pellets. This is a significant distinction as the size of partially visible pellets cannot be correctly estimated with existing size measures and would bias any size estimate. Literature review indicates that other image analysis techniques fail to make this distinction. Statistical classification methods are used to discriminate pellets on the surface of a pile between entirely visible and partially visible pellets. Size estimates of the surface of a pellet pile show that the overlapped particle error is overcome by only estimating the surface size distribution with entirely visible pellets.

1 Introduction

Iron ore pellet's sizes are critical to the efficiency of the blast furnace process in production of steel. Overly coarse pellets effect the blast furnace process negatively, however this effect can be minimized by operating the furnace with different parameters [15]. An on-line system for measurement of the pellet size distribution would improve productivity through fast feedback and efficient control of the blast furnace.

In pellet manufacturing, manual sampling followed by sieving with a square mesh is used for quality control. The manual sampling is performed infrequently and is time-consuming. Fast feedback of the pellet size distribution is desirable.

Thurley [20] present progress on a now completed online imaging and analysis system for non-contact measurement of the size of iron ore green pellets on conveyor belts. A 3D surface data capturing system based on active triangulation is used to collect data. Segmentation of the data is achieved with algorithms based on mathematical morphology for sparse, irregular 3D surface data. It is also shown that sizing of identified pellets gives promising results using the best-fit rectangle [23] measure.

Image analysis techniques promise a quick, inexpensive and non-contact solution to determining the size distribution of a pellet pile. Such techniques capture information of the surface of the pellet pile which is then used to infer the pile size distribution.

However, there are a number of sources of error relevant to surface analysis techniques as follows;

- Segregation and grouping error, more generally known as the brazil nut effect [17], describes the tendency of the pile to separate into groups of similarly sized particles. It is caused by vibration or motion (for example as rocks are transported by truck or conveyor) with large particles being moved to the surface.

- Capturing error [5, 21], describes the varying probability based on size, that a particle will appear on the surface of the pile.

- Partial profile error, describes the fact that only a partial profile of surface particles can be seen making it difficult to estimate size. However, best-fit rectangle [23] has been successfully used as a feature for determining the sieve size of rocks [19] and pellets [20] based on the visible partial profile.

- Overlapped particle error, describes the fact that many particles are only partially visible and a large bias to

the smaller size classes results if they are treated as small entirely visible particles and sized using only their visible profile.

We eliminate both segregation and capturing error from the presented study by comparing the results against the pellets on the surface and not the overall pile size distribution.

Work has been published on size estimation of iron ore pellets that assumes pellets are spherical [2, 3]. However, we have previously shown that spherical fitting is a poor measure of pellet size [1]. More work have been presented on size and shape analysis of rock fragments, and we extend our literature review to include presented work in that field. Comparison of manual sampling and estimates of rock fragment's size using 2D imaging analysis has been published [24, 8]. It is reported by Wang and Stephansson [24] that "a systematic error compared to sieving analysis" is found. With the exception of Thurley [19, 22], 3D surface measurement of rocks has only been applied to segmentation where rocks had little or no overlap [12], or to shape measurements of individual rock fragments [13]. Both Kim et al. [12] and Lee et al [13] propose a mechanical solution to ensure that aggregate particles are not overlapped. However, a mechanical solution is not practical to consider in an operating mine, as it would demand redesign of existing conveyor belt systems.

The presented research extends the work of Thurley [19] and describes an algorithm to overcome overlapped particle error by classifying the pellets on the surface of the pile between entirely visible and partially visible pellets. Once identified, partially visible pellets can be excluded from any surface size estimate.

2 Sample of pellet pile

Mechanical sieving is the accepted industry technique for sizing pellets. A sample of baked iron ore pellets was sieved into 6 size gradings. Each sieve size was painted and color coded to allow manual identification of pellet sizes in mixed pellet piles. The sample was divided into two separate sets. The first set will be used to develop algorithms for visibility and size classification. The second set will be held out during development of the visibility and size classifiers. Thus, the second set will only be used to validate the classifiers performance.

The two sets were loaded onto a laboratory conveyor belt and the stream of pellets was scanned with a 3D imaging system based on laser triangulation. An additional color camera is used to collect color information to overlay on the 3D surface data. A portion of the collected data for the two sets of pellets is shown in figure 1 and 2.

We define the two visibility classes; entirely visible and partially visible. A pellets visibility depend on how much of a pellet is visible from above.
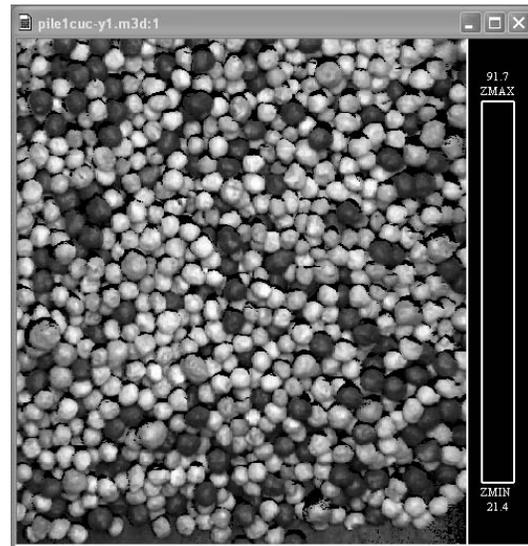

Figure 1. Close up of the first set of pellets captured by 3D imaging system viewed from above. The data is comprised of sparse irregularly spaced 3D coordinate points with overlayed color information.
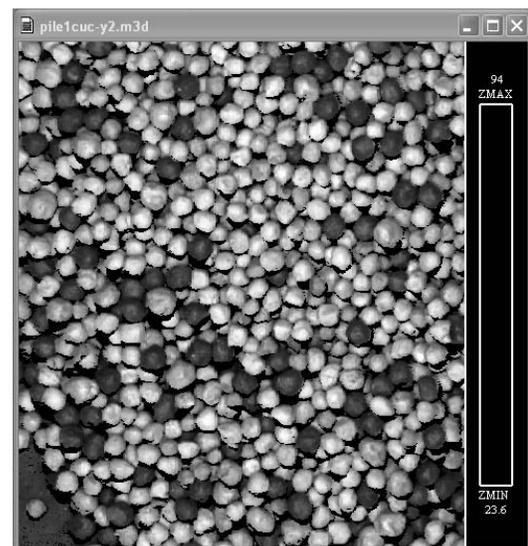

Figure 2. Close up of the second set of pellets.

Paint color *	Size †	All ★	Entirely visible ◇
Orange	6.3	**15.68**	**11.64**
Pink	9	**16.27**	**18.84**
Grey	10	**40.74**	**41.44**
Red	11.2	**21.14**	**20.21**
Yellow	12.5	**4.04**	**5.48**
Green	14	**2.14**	**2.40**

∗ Sieve class color.
† The lower bound of each sieve size increment (mm)
★ Size distribution of all pellets on the surface (%)
◇ Size distribution of only entirely visible pellets on the surface (%)

Table 1. Known size distribution of pellets on the surface for first set. Size distributions of all pellets and only visible pellets are shown.

Paint color *	Size †	All ★	Entirely visible ◇
Orange	6.3	**15.03**	**11.63**
Pink	9	**16.67**	**15.95**
Grey	10	**41.29**	**45.18**
Red	11.2	**17.93**	**19.27**
Yellow	12.5	**6.57**	**6.31**
Green	14	**2.53**	**1.66**

∗ Sieve class color.
† The lower bound of each sieve size increment (mm)
★ Size distribution of all pellets on the surface (%)
◇ Size distribution of only entirely visible pellets on the surface (%)

Table 2. Known size distribution of pellets on the surface for second set. Size distributions of all pellets and visible pellets are shown.

The two sets were manually interrogated to specify each pellet's size and visibility class. The first set has a total number of 842 pellets on the surface of the pile. 292 pellets are labelled as entirely visible and the remaining 550 pellets are partially visible. The sieve size distribution of the pellet pile surface and of entirely visible pellets on the surface for the first set is shown in table 1. The second set has a total number of 792 pellets on the surface of the pile. 301 pellets are labelled as entirely visible and the remaining 491 pellets are partially visible. The sieve size distribution of the complete surface and of the entirely visible pellets on the surface for the second set is shown in table 2.

3   Estimating Pellet Size

As best-fit rectangle [23] has been shown to be a feature with a capacity for discriminating pellets into different sieve sizes [20] we calculate the best-fit rectangle for each pellet in our sample. We visualize the distribution of best-fit rectangle values for entirely visible and partially visible pellets

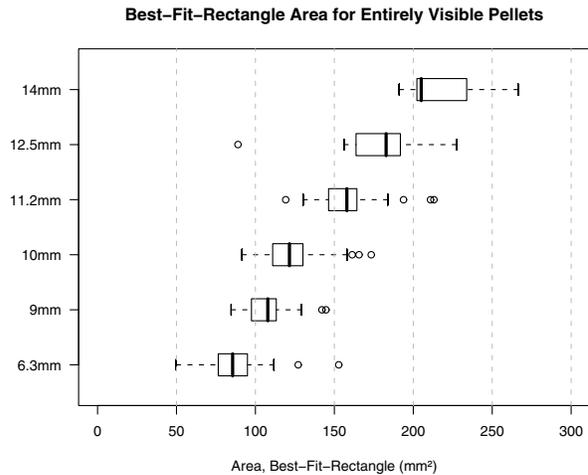**Best–Fit–Rectangle Area for Entirely Visible Pellets**

Figure 3. Distributions of calculated best-fit rectangle for entirely visible pellets. The distribution of the best-fit rectangle values do not overlap

on the surface of a pile using the graphical convention of box-plots in figure 3 and 4.

The central portion of a box-plot [10, 25] contains a rectangular box. In the center of this box is a dashed vertical line, this marks the median value (or 50th percentile) of the data. The left edge of the rectangular box marks the 25th percentile, and the right edge marks the 75th percentile.

In figure 3 box-plots of best-fit rectangle values for the entirely visible pellets are shown. It is clear that an increase of the best-fit rectangle values correlate with an increase of sieve size. It is noticeable that the distributions slightly overlap, although more significantly for the smaller size classes 9 mm and 10 mm. Perfect discrimination into different sieve sized cannot be expected but a majority of the pellets should be possible to discriminate correctly.

In figure 4 box-plots of best-fit rectangle values for the partially visible pellets are shown. As expected, the best-fit rectangle values are shifted to smaller values compared with the values for visible pellets. It is important to notice that these distributions overlap significantly, such that they cannot be discriminated between. This emphasises the unsuitability for sizing partially visible pellets based only on their visible profile.

It is clear that best-fit rectangle values for partially visible pellets cannot be used to estimate sieve size. It is critical to identify these pellets so they can be excluded from any size estimate of pellet piles.
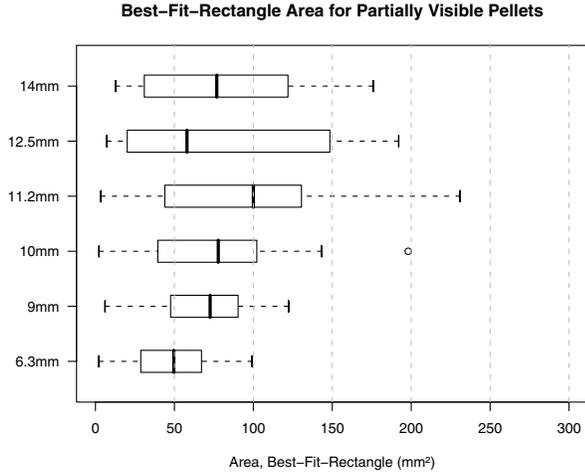
Best–Fit–Rectangle Area for Partially Visible Pellets

Figure 4. Distributions of calculated best-fit rectangle for partially visible pellets. The distribution of the best-fit rectangle values overlap.

4 Classification

Generally all classifiers try to predict a *response*, here denoted y, from a set of *feature values*, here denoted $\mathbf{x}$. Detailed information can be found in the books [7] and [11].

In this research, we propose a method to overcome overlapped particle error when estimating the surface size distribution of a pellet pile. Logistic regression is used to classify pellets on the surface of a pile between two visibility classes; entirely visible and partially visible. As partially visible pellets cannot be sized correctly with the calculated best-fit rectangle area, partially visible pellets must be excluded from any estimate of the surface size distribution of a pellet pile.

We then apply logistic regression to classify visible pellets into different sieve size classes based on the calculated best-fit rectangle area.

4.1 Feature Extraction

In image analysis, shape analysis is a common approach to describe and classify specific objects, or regions, in an image.

2D shape features have been used to detect broad-leaved weeds in cereal crops [16], to allow a rover to classify the shape and other geologic characteristics of rocks [9], to investigate the suitability of an imaging system to measure shape of particles [4] and for detection and classification of

rocks [18]. A 3D feature called visibility ratio have been used to classify the visibility of rocks in piles [19].

In this work we extract 25 different features to describe each pellet. These are a collection of shape features used by the above authors. There is no room for a description of all features here, a brief description of the selected features will be given later in this text.

4.2 Classification methods

The distribution of feature values in a data set is important to investigate in order to choose the right classification method. Inspection of our data set shows that the feature values are not multivariate normally distributed. The type of the response value also need to be considered when classification method is chosen. The response variable's type is binary for the visibility classification as the visibility class of a pellet is either entirely visible or partially visible. The response variable's type for size classification is ordinal as a pellet's size class range from sieve size 6.3 mm to 14 mm.

Johnson [11] suggests to use logistic regression as a classification method when the features values distribution are not multivariate normal. A rigorous description of logistic regression can be found in An Introduction to Generalized Linear Models [6].

Logistic regression can be used when the response variable are binary, ordinal or nominal. In the case when the response variable can only take two values, the method is called binary logistic regression. The form of the logistic regression model is shown in equation 1 where y is the response variable, $\mathbf{x}$ is a feature vector, β_0 is a constant and β_1 is a vector of parameters. $P(y = 1|\mathbf{x})$ denotes the probability that $y = 1$ given the observed feature vector $\mathbf{x}$. The model, or more specifically, β_0 and β_1 is fit to the known data via the use of maximum likelihood estimation.

$$P(y = 1|\mathbf{x}) = \frac{e^{\beta_0 + \beta_1'\mathbf{x}}}{1 + e^{\beta_0 + \beta_1'\mathbf{x}}} \qquad (1)$$

For response variables that have a natural order, the order can be used to form an expression for the cumulative probability using ordinal logistic regression. In equation 2 the cumulative probability is shown, where the possible response values is $i = 1, 2, ..., J$ and J is the number of possible response values.

$$P(y \leq i) = \sum_{j=1}^{i} P(y = j|\mathbf{x}) = \frac{e^{\beta_{0j} + \beta'\mathbf{x}}}{1 + e^{\beta_{0j} + \beta'\mathbf{x}}} \qquad (2)$$

From equation 2 the probability for each category given a feature set is easily derived knowing that $P(y \leq J) = 1$.

Logically, the response of the classifiers are $y = j$ where $P(y = j|\mathbf{x}) > P(y = k|\mathbf{x})1$ for all $k \neq j$.

4.3 Feature Selection

As stated before, 25 features are extracted to describe each pellet on the surface of the pile. Some features are strongly correlated and some do not contribute with information that can be used to discriminate between the two visibility classes. An important step when designing a classifier is to select a set of features that can be used to discriminate between desired classes efficiently. We use backward elimination to find a set of features that are statistically significant for discriminating between different size classes of pellets.

Backward elimination of features is an iterative technique that includes all features in the model as an initial step. The technique tests whether there are features in the model that are statistically insignificant and remove the least significant one. It is important to not remove multiple features in each iteration even though they are determined to be statistically insignificant. Features may be insignificant in combination with other features but significant when those features are removed. The iterative process stops when a set of features is obtained where all features are found to be statistically significant.

To test whether a feature is statistically insignificant the parameters β_{0j} and β are fit to the data using maximum likelihood estimation, then Wald statistics for each feature are calculated. Wald statistics are calculated by equation 3 where Z is Walds chi-square value, b is the parameter estimated for a feature and σ_b is the estimated variance of b. Z is then compared against a chi-square distribution to obtain a p-value that will indicate whether a feature is statistically insignificant. If the p-value for a feature is larger than the predetermined significance level, then the feature is deemed insignificant and may be removed. In every iteration the feature with the largest p-value above the predetermined significance level is removed.

$$Z = \frac{b^2}{\sigma_b} \qquad (3)$$

Using backward elimination with a significance level of 2%, 4 statistically significant features are selected for discriminating between entirely visible and partially visible pellets. The final set of features are:

- **Equivalent area diameter** [23] is the diameter of a circle with equal area as the region of interest. The equivalent area diameter is calculated by equation 4.

$$ED = 2 * \sqrt{\frac{Area}{\pi}} \qquad (4)$$

- **Visibility ratio** [19] is a boundary following algorithm that accommodates *sparse, irregularly spaced* 3D co-

		Predicted	
		Entirely visible	Partially visible
Known	Entirely visible	**89.7**	10.3
	Partially visible	3.87	**96.13**

Table 3. Confusion matrix that show how entirely visible and partially visible pellets in the second pile are classified.

ordinate data to allow the determination of entirely visible and partially visible rocks.

- **Minor axis** [14] is the length of the minor axis of the ellipse that has the same normalized second central moments as the region.

- **Major axis** [14] is the length of the major axis of the ellipse that has the same normalized second central moments as the region.

4.4 Validation

How well the visibility and sizing classifiers perform are validated using the holdout method. The holdout method is a technique where a classifier is developed on a specific training set. A test set, separate from the training set, is used to estimate how well the classifier performs on new data. This method gives an unbiased estimate of classifiers performance. As our data consist of two separate piles of pellets collected in the same conditions, we use the first pile as the training set and the second pile as the test set.

5 Validation of Visibility Classification

In table 3 a confusion matrix is presented for the visibility classification results of the second pile. Binary logistic regression with a feature set composed by effective diameter, major axis, minor axis and visibility ratio is used. 89.7 % of the visible pellets and 96.13% of the partially visible pellets are classified correctly.

6 Overcoming Overlapped Particle Error

To show how the identification of partially visible pellets may overcome overlapped particle error, ordinal logistic regression is used to classify each entirely visible pellet into a sieve size class. Best-fit rectangle area is used by itself to discriminate between the different size classes.

The sizing classification results for the entirely visible pellets can be seen in table 4. The confusion matrix shows the classification accuracy for pellets of size class 6.3 mm,

		Predicted size class (mm)					
	(mm)	6.3	9	10	11.2	12.5	14
Known size class	6.3	**84.21**	15.78	0	0	0	0
	9	18.18	**24.64**	57.57	0	0	0
	10	0	12.77	**82.98**	4.26	0	0
	11.2	0	0	19.05	**73.81**	7.14	0
	12.5	0	0	7.14	21.43	**14.28**	57.14
	14	0	0	0	0	0	**100**

Table 4. Confusion matrix (percentages) that show the sizing classification results for entirely visible pellets. The table show how pellets of each size class is classified.
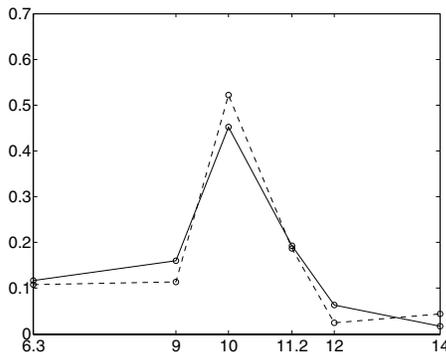

Figure 5. Known and estimated surface size distribution for the entirely visible pellets. The solid line is the known and the dashed line is the estimated surface size distribution of entirely visible pellets.

10 mm, 11.2 mm and 14 mm is above 73 %. The classification accuracy is low for the two size classes 9 mm and 12.5 mm. However, we note that pellets that are misclassified are classified to a size class close to the known size class.

Even though perfect classification of pellet's size class is not achieved, an estimate of the surface size distribution is achieved for the entirely visible pellets. In figure 5 known and estimated surface size distribution is shown for the entirely visible pellets on the surface of the pile. The dashed line, which is the estimated surface size distribution, follow the solid line, which is the known surface size distribution.

7 Conclusion

Visibility classification of pellets in a pile have been presented to overcome overlapped particle error. Pellets were collected and manually sieved into different sieve size classes. The pellets were mixed in a pile and scanned with a 3D camera system. We define two visibility classes; en-

tirely visible and partially visible. This is a significant distinction as partially visible pellet's size cannot be correctly estimated with existing size measures and would bias any size estimate. We overcome overlapped particle error by only estimating the surface size distribution with entirely visible pellets. Binary logistic regression is used with 4 optimal features to describe a pellet selected from a total of 25 features. Holdout method is used to estimate the visibility classifier's accuracy to predict 89.7 % of the entirely visible and 96.13 % of the partially visible pellets correctly. It is shown that the surface size distribution of the visible pellets can be estimated correctly using best-fit rectangle.

8 Acknowledgment

References

[1] T. Andersson, M. Thurley, and O. Marklund. Pellet size estimation using spherical fitting. *Proceedings of the IEEE Instrumentation and Measurement Technology Conference*, pages 1–5, May 2007.

[2] M. Blomquist and A. Wernerson. Range camera on conveyor belts: estimating size distribution and systematic errors due to occlusion. *Proceedings of the SPIE - The International Society for Optical Engineering*, 3835:118–126, Sept 1999.

[3] A. Bouajila, M. Bourassa, J.-A. Boivin, G. Ouellet, and T. I. Martinovic. On-line non-intrusive measurement of greeen pellet diameter. *Ironmaking Conference Proceedings*, pages 1009–1020, 1998.

[4] R. Carter and Y. Yan. Measurement of particle shape using digital imaging techniques. *Journal of Physics: Conference Series*, 15:177–182, 2005.

[5] R. Chavez, N. Cheimanoff, and J. Schleifer. Sampling problems during grain size distribution measurements. *Proceedings of the Fifth International Symposium on Rock Fragmentation by Blasting - FRAGBLAST 5*, pages 245–252, Aug 1996.

[6] A. J. Dobson. *An introduction to generalized linear models*. Chapman & Hall/CRC, 2nd ed. edition, 2002. ISBN: 1-58488-165-8.

[7] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. Wiley, 2. ed. edition, 2001. ISBN: 0-471-05669-3.

[8] J. Fernlund. The effect of particle form on sieve analysis: a test by image analysis. *Engineering Geology*, 50:111–124, 1998.

[9] J. Fox, R. Catao, and R. Anderson. Onboard autonomous rock shape analysis for mars rovers. *Proceedings of the IEEE Aerospace Conference*, 5, 2002.

[10] R. Ihaka and R. Gentleman. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3):299–314, 1996. http://www.r-project.org/.

[11] D. E. Johnson. *Applied Multivariate Methods for Data Analysts*. Duxbury Press, 1998. ISBN: 0-534-23796-7.

[12] H. Kim, A. Rauch, and C. Haas. Automated quality assessment of stone aggregates based on laser imaging and a neural network. *Journal of Computing in Civil Engineering*, pages 58–64, January 2004.

[13] J. Lee, M. Smith, L. Smith, and P. Midha. A mathematical morphology approach to image based 3d particle shape analysis. In *Machine Vision and Applications*, volume 16(5), pages 282–288. Springer–Verlag, 2005.

[14] Matlab®. Matlab image processing toolbox 5, users guide. page 560, 2007. Release 2007a.

[15] L. S. Ökvist, A. Dahlstedt, and M. Hallin. The effect on blast furnace process of changed pellet size as a result of segregation in raw material handling. *Ironmaking Conference Proceedings*, pages 167–178, 2001.

[16] A. Perez, F. Lopez, J. Benlloch, and S. Christensen. Colour and shape analysis techniques for weed detection in cereal fields. In *Proceedings of 1st European Conference for Information Technology in Agriculture*. Elsevier, February 2000.

[17] A. Rosato, K. J. Strandburg, F. Prinz, and R. H. Swendsen. Why the brazil nuts are on top: Size segregation of particulate matter by shaking. *Physical Review Letter*, 58:1038 – 1040, 1987.

[18] D. Thompson, S. Niekum, T. Smith, and D. Wettergreen. Automatic detection and classification of features of geologic interest. *Proceedings of the IEEE Aerospace Conference*.

[19] M. J. Thurley. *Three Dimensional Data Analysis for the Separation and Sizing of Rock Piles in Mining*. PhD thesis, Monash University, December 2002. Download from http://image3d6.eng.monash.edu.au/thesis.html.

[20] M. J. Thurley. On-line 3d surface measurement of iron ore green pellets. In *Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation*. IEEE, November 2006.

[21] M. J. Thurley and K. C. Ng. Modelling the relationship between the surface fragments and the pile size distribution in laboratory rock piles. *In submission*.

[22] M. J. Thurley and K. C. Ng. Identifying, visualizing, and comparing regions in irregularly spaced 3d surface data. *Computer Vision and Image Understanding*, 98(2):239–270, February 2005.

[23] W. Wang. Image analysis of particles by modified ferret method - best-fit rectangle. *Powder Technolgy*, 165(1):1–10, Jun 2006.

[24] W. Wang and O. Stephansson. Comparison between sieving and image analysis of aggregates. pages 141–148.

[25] Wikipedia. Box plot. http://en.wikipedia.org/wiki/Boxplot, Jul 2007.