WATERSHED-BASED SEGMENTATION OF ROCK SCENES AND PROXIMITY-BASED CLASSIFICATION OF WATERSHED REGIONS UNDER UNCONTROLLED LIGHTING.

S.G. Mkwelo*, F. Nicolls and G. De Jager*****

** Department of electrical engineering, University of Cape Town, Rondebosch, 7700*
*** Department of electrical engineering, University of Cape Town, Rondebosch, 7700*
**** Department of electrical engineering, University of Cape Town, Rondebosch, 7700*

Abstract: A watershed-based segmentation approach that uses iterative bilateral smoothing with an adapting photometric similarity parameter for pre-filtering is adopted for the segmentation of rock scenes. However, the resultant segmented images also contain non-rock watershed regions that are not desired for measuring rock sizes. A proximity-based classifier is applied for the removal of the latter using features that can be divided into rock shape, edge strength and region intensity characteristics. Subset feature selection based on Thornton's separability index is used to remove redundant and irrelevant features. We achieve final classification rates of 89.91% using the simple k-nearest neighbor (KNN) classifier.

Key words: Bilateral filtering, watershed segmentation, proximity-based classification, rock size distribution.

1. INTRODUCTION

The mineral processing industry requires an instrument that can segment a scene of rocks on a conveyor belt in order to facilitate accurate measurement of rock size distributions. The Machine Vision system under construction has possible applications in control and optimization of milling machines. We define a rock size to be the projected surface area of a rock due to the constraint imposed by the 2D nature of an image. Based on this assumption, a rock scene can be accurately segmented by applying a robust edge detector to find rock boundaries. A wide range of edge detectors have been implemented and improved but they are not suitable, as they require cleaning and joining of the detected pixels.

A method of choice for many image segmentation applications is the watershed transform [1]. In this approach, an image intensity map is viewed as a topographical landscape where intensity minima are catchment basins and the ridges are the watersheds [2]. The objective is to search for the watersheds by region growing, growing catchment basins from a predefined set of local minima until each pixel in the image belongs to one of the labeled catchment basins. Formal definitions and implementations can be found in [1]. A transformation of an original image to a gradient image followed by an application of the watershed transform should ideally find rock edges in an image. The general shortcomings of this approach are over-segmentation and sensitivity to noise. These can be overcome by careful implementation of a pre-filtering scheme and a procedure for selecting a set of predefined local minima.

Linear pre-filtering methods such as the Gaussian filter are not suitable for this application since they blur surfaces and edges equally. Instead, edge-preserving techniques such as anisotropic diffusion and bilateral filtering are used. These approaches smooth across surfaces while treating edges as outliers and thus preserving them [3,4]. Anisotropic diffusion is an iterative procedure based on a nonlinear anisotropic version of the heat diffusion equations proposed by Perona and Malik [3]. It is highly iterative and this is not a desirable property for real-time applications. On the other hand a bilateral filter is non-iterative and the filtering is based on spatial closeness and photometric similarity of pixels [4]. The key issue with these approaches is deciding how significant an edge should be for it to be considered an outlier [3]. Work has been done in [3] to address this issue.

However, these approaches fail when there are cracks across rock surfaces. As a result the watershed lines follow these cracks giving rise to a rock splitting phenomenon. In this work rock splitting is overcome using a combination of iterative bilateral filtering with an adapting photometric similarity parameter followed by a hierarchical analysis of a stack of thresholded watershed images. In this framework, watershed lines that are likely to represent rock boundaries are retained and the rest are discarded. A proximity-based classifier is then used to eliminate spurious regions.

This paper is organized as follows. In the next section we describe the pre-processing that is involved. In section 3, the overall watershed based segmentation method is described. This is followed by the classification of watershed regions in section 4. Results are presented in section 5. Finally conclusions are drawn in section 6.

2. PREPROCESSING

In this section we describe the early image processing operations that are used to achieve rock location estimation. This stage is necessary for predefining a set of minima for the watershed transform. It is divided into adaptive thresholding, distance transform and peak detection by h-domes operations.

2.1 Adaptive thresholding

An adaptive threshold is computed as the mean of the neighborhood pixel intensities for each pixel in the original image. The key parameter is the neighborhood size, which is critical for the detection of a limited range of rock sizes. In this work, we use a 2-window approach where the smaller window of size 25x25 is used for detecting small rocks and the larger one of size 95x95 is used for detecting larger rocks in a 240x240 image. A logical or operation is then applied to combine the results. The resultant output together with the original image is shown below.

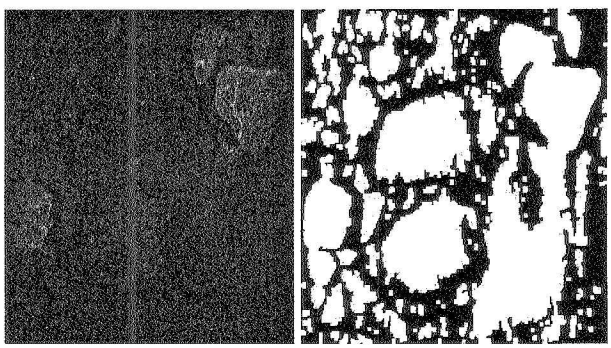

Figure 1: The original image and its' thresholded version

2.2 Applying a distance transform (DT)

The result of the previous operation usually contains connected white areas that are supposed to be separated areas. To resolve this, we apply a distance transform (DT) to the inverted version of the thresholded image so that the locations of the local maxima of the DT output are the approximate blob centers. Rock locations can be estimated by extracting the peaks of distance transform (DT).

2.2 DT peak extraction by h-domes

At this stage we aim to extract the peak locations of the distance transform. This is achieved using the h-domes operation, which is based on greyscale reconstruction by dilation. Greyscale reconstruction literature can be found in [5]. The sequence of operations commences with a vertical down shift of the DT by a constant h, followed by a grey-scale reconstruction by dilation of the original DT using the shifted version. The drawback of this procedure is its sensitivity to the h-parameter. If set too high, small blobs are missed and if too low, large and non-circular blobs are split. One approach to solving this is to compute the centroids of white areas for the correction of incorrect splitting during segmentation. Markers obtained by both methods are shown below.

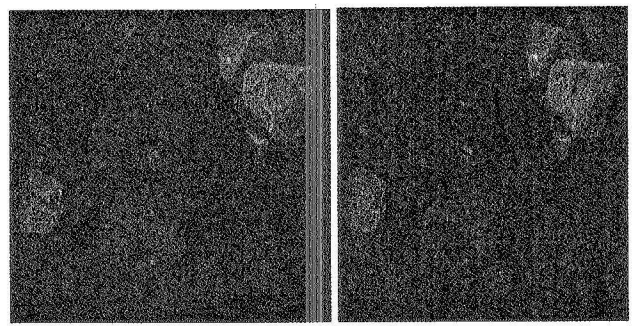

Figure 2: The blob centroids and the distance transform peak locations

3. THE WATERSHED-BASED SEGMENTATION OF ROCK SCENES

The watershed-based segmentation approach consists of an iterative bilateral pre-filter with an adapting photometric similarity parameter, a watershed transform for finding rock edges and a scheme for finding watershed lines that are most likely to be rock edges. The overall structure is shown in figure 3. In this structure, the input signal is the original image, F1 denotes the first filtering operation, G1 is the gradient operation on the first filtered image, DM is the marker image from the distance transform, CM is the marker image containing centroids of white areas and W1 is the first watershed transformation.

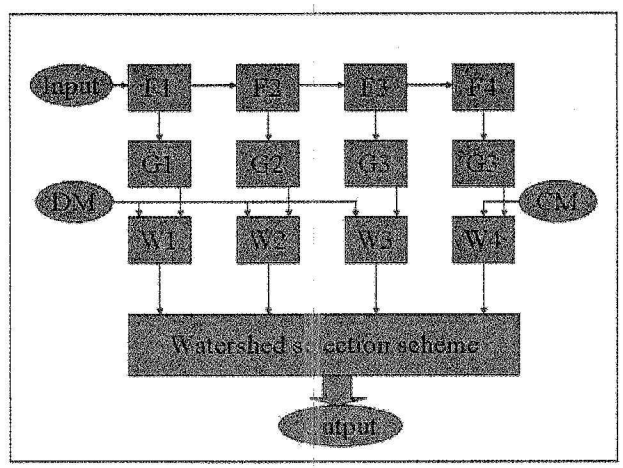

Figure 3: Segmentation algorithm structure

3.1 Bilateral filtering

Theory: In this application, a filter that preserves edges is required as these are indicators of rock boundaries. The Bilateral filter obeys this requirement. It performs a combination of both domain and range filtering

operations based on spatial closeness and photometric similarity respectively [4]. The bilateral filter for grayscale images as defined by Tomasi and Manduchi has the form:

$$h(\mathbf{x}) = k^{-1}(\mathbf{x}) \iint f(\xi)c(\xi, \mathbf{x})s(f(\mathbf{x}), f(\xi))d\xi, \quad (1)$$

Where the normalization term is

$$k(\mathbf{x}) = \iint c(\xi, \mathbf{x})s(f(\xi), f(\mathbf{x}))d\xi. \quad (2)$$

A pixel value at $\mathbf{x}$ is replaced with a linear combination of similar and nearby pixels. The terms $c(\xi, \mathbf{x})$ and $s(f(\xi), f(\mathbf{x}))$ are the closeness and photometric similarity functions of the Euclidean distance between their arguments respectively. Both functions should generally decrease with increasing distance. A common case is where these functions are defined as Gaussian functions of the Euclidean distances.

Gaussian bilateral filtering: In this version, the spatial closeness and photometric similarity functions are defined as Gaussian functions of their arguments as shown below.

$$c(\xi, \mathbf{x}) = e^{-1/2(d(\xi,\mathbf{x})/\sigma_d)^2}. \quad (3)$$

Where:

$$d(\xi, \mathbf{x}) = \|\xi - \mathbf{x}\| \quad (4)$$

is the Euclidean distance between ξ and $\mathbf{x}$. The photometric similarity is defined as:

$$s(f(\xi), f(\mathbf{x})) = e^{-1/2(\delta(f(\xi),f(\mathbf{x}))/\sigma_r)^2}. \quad (5)$$

Where:

$$\delta(f(\xi), f(\mathbf{x})) = \|f(\xi) - f(\mathbf{x})\| \quad (6)$$

is the absolute difference of the intensity values f(ξ) and f($\mathbf{x}$). This implementation of the bilateral filter requires prior values of the space parameter σ_d and the similarity or range scale parameter σ_r. The space parameter has little effect on the preservation of edges, it only imposes a closeness constraint so that pixels far away from the kernel center have little influence on the kernel weighted mean. On the other hand, the value of the photometric similarity parameter is critical and can be estimated using the global gradient variation as defined in [3] for the anisotropic diffusion technique.

$$\sigma_r = 1.4826 \times median(\|\nabla I - median_I(\|\nabla I\|)\|) \quad (7)$$

Where: $\|\nabla I\|$ is the gradient magnitude of the image I. A more accurate estimate of this parameter can be achieved by computing the local gradient variation in predefined neighborhoods and interpolating to cover the whole image. A criterion for an optimal similarity parameter is rock edge preservation and surface crack smoothing. This requirement cannot be met using neither a single filtering operation nor iterated filtering where a single value for σ_r is used because there is a trade-off on the magnitude of σ_r. As a cure for this, an iterative bilateral filter with an adapting σ_r value is proposed.

Iterative bilateral filtering with an adapting σ_r parameter: In this approach, the image is effectively filtered n-times with the range scale parameter σ_r estimated each time using equation 7 and weighted accordingly. The weights increase monotonically from the first to the last filtering operation so that the resultant n-images exhibit fewer cracks as one moves from the first to the last filtered image. Multi-scalar techniques are usually performed in the domain direction to facilitate the detection of a wide range of object sizes. In this application the domain scale is kept constant and the range scale parameter is varied to detect edges of various strengths. This scheme forms the first stage of our watershed-based segmentation approach as shown in figure 3.

3.2 The gradient watershed transform

A gradient watershed transform can be classified as an edge detector because it locates regions of high gradient strengths given a gradient image as the input. Its main drawback is over-segmentation, which is usually due to the inaccurate determination of markers. In this work, we use the outputs of the pre-processing algorithm as the set of markers as shown in figure 3. A parallel watershed scheme is then applied to the filtered images as shown in figure 3.

Multiple watersheds analysis: This approach considers a collection of n-binary watershed images as containing a population of edge samples that is likely to contain the desired combination of edge samples. A summation of the n-binary watershed images followed by a multi-level thresholding procedure is firstly executed. The n thresholds are integers with the minimum threshold of unity and the maximum n. The output is a stack of n

binary watershed images with the bottom level image containing all the watersheds from the n-images and the top image containing the watersheds that are most likely to be rock edges but have poor pixel connectivity. The underlying assumption is that a watershed line is likely to be a rock edge if it survives most of the thresholding. The watershed image from the centroid markers has highly merged regions and is placed at the top of the stack as shown in figure 3.

The evaluation algorithm collects regions starting from the highly merged regions at the top of the stack to the highly split regions at the bottom. The selection criterion is:

$$(\frac{A_{blob}}{A_{region}} \geq \phi) \cap (C \geq \rho) \quad . \tag{8}$$

Where A_{blob} is the area of the corresponding white area in the thresholded image, A_{region} is the area of the watershed region of interest, ρ is a threshold value that ranges between 0 and 0.75, ϕ is a threshold value that ranges between 0.5 and 0.8 and C is the circularity of the watershed region of interest which is calculated as:

$$C = \frac{4\pi A_{region}}{P^2} . \tag{9}$$

Where P is the perimeter of the watershed region. The final output of the segmentation algorithm is shown in figure 4.

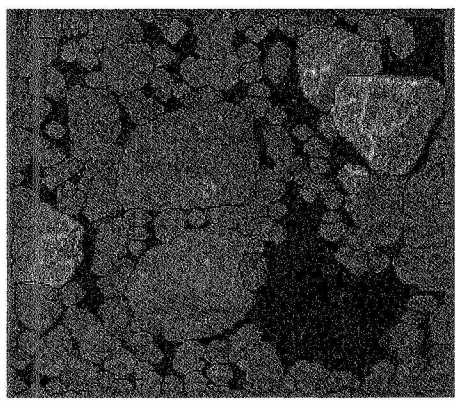

Figure 4: Segmentation output

4. CLASSIFICATION

As can be seen in figure 4 that the segmentation output contains spurious regions in addition to rock regions. The purpose of a classifier in this context is to discriminate between rock and non-rock regions. In this section we present the adopted methodology for data collection, feature descriptions, feature subset selection results, a brief description of the k-nearest neighbor (KNN) classifier and its application to the problem.

4.1 Methodology

A data set consisting of 20 images is assembled. The images were captured from two mineral processing plants under different lighting conditions. The type of ore that is handled by the plants is also different in terms of color and texture. After watershed-based segmentation, a new set of images that has 2415 regions or "potential rocks", an average of 120 regions per image and the number of regions per image ranges between 86 and 146 regions, is formed. The data set is partitioned into training, validation and testing sets with proportions as shown in table 1. The training set is a mixture of images obtained from the two mineral plants in equal proportions and, the validation and test-sets are obtained from one plant. Therefore the system is expected to generalize well to images captured from these two mineral processing plants. The labeling of examples is performed manually using the author's discretion and thus an element of human error is expected.

Table 1: Data partitioning

Partition	Images	Regions	Fraction of actual rocks
Training	10	1117	22.74%
Validation	5	644	13.35%
Testing	5	654	12.69%
Total	20	2415	17.52%

4.2 Feature extraction

Eleven features are measured and can be broadly divided into rock shape, edge and gray level characteristics.

Centroid to boundary distance variance feature: is measured by firstly finding the centroid of the region $\mathbf{c}$ and its boundary pixel coordinates $\mathbf{b}$. This measure is computed as the variance of the distances from $\mathbf{c}$ to each of the $\mathbf{b}_i$ pixels.

Proportion of dark interior pixels feature: is computed as the ratio of the number of detected dark interior pixels of the region to the total area of the region.

Proportion of dark boundary pixels feature: is computed as the ratio of the number of dark boundary pixels to the region perimeter.

Proportion of thresholded area to watershed region area feature: is computed as the ratio of the area of the region in the adaptively thresholded image to the area of the corresponding watershed region of interest.

Average interior gray level feature: is computed as the ratio of the total gray level on a small disk around the region centroid to the area of the disk.

Average boundary gray level feature: is computed as the ratio of the total boundary gray level to the perimeter of the region.

Average interior gray level gradient feature: is computed as the ratio of the gray level gradient on a small disk around the region centroid to the area of the disk.

Average boundary gray level gradient feature: is computed as the ratio of the gray level gradient on the boundary to the watershed region perimeter.

Boundary and interior gray level absolute difference: is computed as the absolute difference of the boundary and interior gray level features.

Boundary and interior gradient absolute difference: is computed as the absolute difference of the boundary and interior gradient features.

Interior gray level variance: is computed as the variance of the gray levels inside the region.

4.3 Feature subset selection

Proximity based classifiers are sensitive to irrelevant and redundant features [6] which impair the separability of the data in the feature space. Feature subset selection procedures attempt to remove such features and thus improve the separability of the data. In this work, we use Thornton's separability index as an optimization objective function such that the feature set with the maximum separability index is sought. A separability index of unity represents the case where the classes are separate with no overlap while a zero index represents total mixing of the classes. An evaluation of the separability index function is carried out for all the 2047 possible combinations of features using equation 10:

$$si = \frac{\sum_{i=1}^{n} (f(x_i) + f(x_i') + 1) \bmod 2}{n}, \qquad (10)$$

where: x_i' is the nearest neighbor of x_i and n is the number of training examples or points. After these evaluations, the feature combination with the maximum separability index is selected. Table 2 shows the maximum separability index when all the features are present and after the optimal combination was selected on the training set. A set of 6 of the 11 features consisting of the centroid to boundary distance variance, proportion of dark interior pixels, average interior gray level, average boundary gray level, difference of interior and boundary gray levels, and the difference between interior and boundary average gradients are retained as the optimal feature set.

Table 2: Table of separability indices.

All features	Optimal features
78.25%	82.45%

4.4 Classifier training

This section provides a brief description of the k-nearest neighbor classifier and the training results.

The k-nearest neighbor (KNN): A query point is assigned a label of the majority of its k-nearest training points. The majority vote is achieved by averaging the labels of the k-nearest training points so that the negative effect of erroneous points in the data is cancelled. Euclidean distance is commonly used for quantifying distance in an N-dimensional space. The parameter k is an odd integer and is selected by cross validation on independent data.

Finding the optimal k value: A range of k-values between 1 and 25 are pre-selected and for each value the KNN algorithm is executed for the labeling of the validation data set whose labels are pre-determined manually for comparison purposes. The percentage accuracy is recorded as the percentage of points that are correctly labeled and is plotted in figure 5. The k-value with the highest accuracy is selected as shown in table 3.

Table 3: The optimal k-value

k-value	Percentage accuracy
13	90%

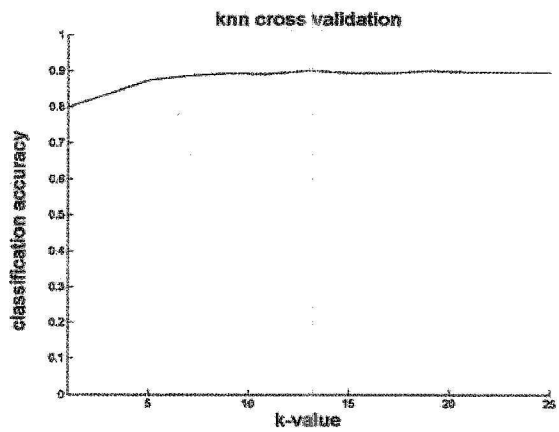

Figure 5: Training results

5. RESULTS

5.1 Testing for generalization

The performance of the KNN is evaluated on independent test data to test its generalization to new unseen data. The test set consists of 654 feature vectors from 5 images. Classification rates of 89.9 percent are achieved using a k-nearest neighbor with the k-value of 13 as shown in table 4. The visual results of classifying the regions on the 5 images are shown in figure 6. A visual comparison of the manually classified regions and the KNN classified regions shows that the classifier has generalized well.

Table 4: Generalization accuracy

k-value	Classification accuracy
13	89.91%

Other classifiers in the form of the Probabilistic Neural Network (PNN) [8], the Kernel Adatron Support Vector Machine (SVM) [7] and the Regularized Least Squares Classifier (RLSC) [9] can also be applied on the test-set. In general, these techniques assign a Gaussian kernel to each training point so that test points that are closer to that training point have higher activation values than those that are further away. A weighted sum of labeled activations is performed to determine the decision function value for that test point. Finally, a threshold (usually set at zero) is applied to determine the class label of the test point.

These techniques differ in how the weights of decision function are determined and this subsequently affects the smoothness of the decision boundaries that are formed in the feature space. The details of these differences can be found in [10]. From table 5, it can be seen that the regularized least squares classifier (RLSC) outperforms all the other classifiers on this data set.

Table 5: Performances of other classifiers

PNN	SVM	RLSC
91.59%	90.07%	92.35%

6. CONCLUSIONS

Based on the above findings and results, the following conclusions can be drawn.

- The rock scene segmentation algorithm does trace the rock edges as required. However, the problem of quantifying its performance is not tackled.
- The KNN classification results show that the selected set of features is effective and confirms 89.9 percent of the target labels.
- A comparative evaluation of various classifiers shows that RLSC outperforms the other approaches on the test set.

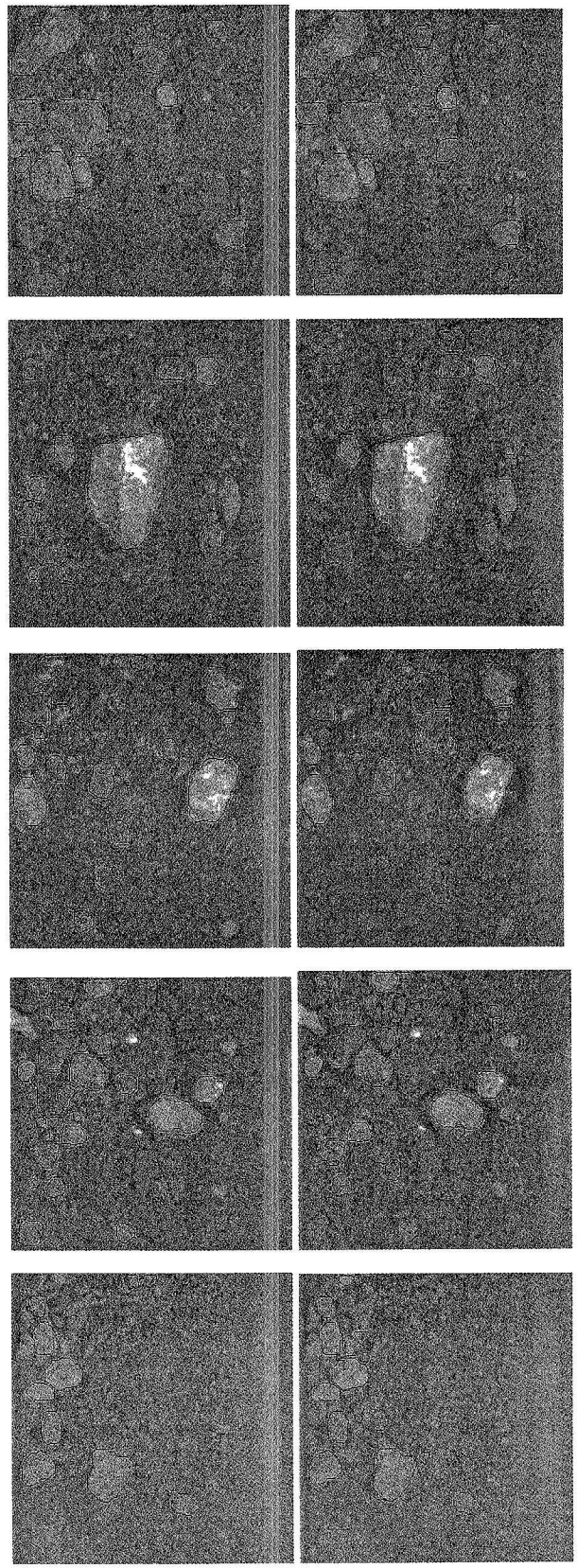

Figure 6: Test results showing manually classified regions on the left and KNN classified regions on the right.

7. ACKNOWLEDGEMENTS

I am grateful to Amplats, MPRU, UCT and the NRF for financial and material support.

8. REFERENCES

[1] B. T. M. Roerdink and A. Meijster, "The Watershed Transform: Definitions, Algorithms and Parallelization Strategies", *Fundamenta Informaticae*, IOS Press, 41(2000), 2000.

[2] A. S. Wright and S. T Acton, "Watershed Pyramids for edge detection", *IEEE*,1997.

[3] M. J Black, "Robust Anisotropic Diffusion", *IEEE Trans. on Image Proc.*, 7(3), 1998.

[4] Tomasi, C. and Manduchi, R., "Bilateral Filtering for gray and color images", *Proc. of the IEEE International Conference on Computer Vision*, 1998.

[5] S. Beucher and F. Meyer, "The morphological approach of segmentation: the watershed transformation". In E. Dougherty, editor, *Mathematical Morphology in Image Processing*, pages 433 to 481. Marcel Dekker, New York, 1992.

[6] J.R Greene, "Feature Subset using Thornton's Separability index and its applicability to a number of sparse proximity based classifiers", *Proceedings of the 12th Annual Symposium of the Pattern Recognition Association of South Africa*, 2001.

[7] D.J. Mashao, "Comparing SVM and GMM on parametric feature sets", *Proceedings of the 14th Annual Symposium of the Pattern Recognition Association of South Africa*, 2003.

[8] D.F. Specht, "Probabilistic neural networks and the Polynomial adaline as complementary techniques for classification", *IEEE Trans. on neural netoworks*, March 1990.

[9] T. Poggio and S. Smale, "The mathematics of learning: Dealing with data", *Notices of the AMS*, 50(5), 2003.

[10] S.G Mkwelo, "A machine vision-based approach to measuring the size distribution of rocks on a conveyor belt", *An MSc Thesis in the department of electrical engineering*, University of Cape Town, 2004.