

https://dspace.org/

Automatic indexing of South African rock art images

Purshotam, Amrit

2015
http://hdl.handle.net/10539/18583

Downloaded from a DSpace repository. https://dspace.org/

Automatic Indexing of South African Rock Art Images

Amrit Purshotam



A dissertation submitted for the degree of
Master of Science

School of Computer Science
University of the Witwatersrand
Private Bag 3, Wits 2050, Johannesburg, South Africa

2015

Abstract

Rock art is the archaeological term used to describe pre-historic artworks placed on natural stone and as one of the earliest known traces of modern human creativity, it is a major component of world history and human heritage. Archival records and the art itself, however, are rapidly decaying requiring the need to preserve them for future generations and humanity as a whole. In line with this, the Rock Art Research Institute digitised their collections of photographs and historical records of the rock art in southern Africa. This has resulted in the South African Rock Art Digital Archive, a collection of over 275 000 images of paintings depicting a wide variety of objects such as humans and animals. The problem with this archive, however, is that most of the images are not labelled with the objects that appear in them. Manual labelling is infeasible due to the size of the archive but rock art researchers require this information to perform text-based search queries. Therefore, in this research, we present the combination of the Viola Jones object detection framework and a Support Vector Machine to automatically classify rock art objects. To test it, we have created and assessed the performance of classifiers for eland, elephant, human, and rhebuck rock art objects. We have performed the experiments using five-fold cross-validation and found the results to be promising considering the variation and deterioration of the paintings.

Declaration

I, Amrit Purshotam, declare that this dissertation is my own, unaided work unless explicitly referenced. It is being submitted for the degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other university.

....

Signature *Date*

Acknowledgements

I would like to thank my primary supervisor, Dr Hima Vadapalli, and co-supervisor, Professor Turgay Celik, for sharing their knowledge of image processing and machine learning as well as their ideas to tackle this research. I also appreciate their availability and assistance whenever I needed them throughout the two years of my degree.

I would also like to express my gratitude to Professor David Pearce for sharing his expertise in the field of southern African rock art and for his kind suggestions and advice to improve my articles. Also for his expertise in southern African rock art and for allowing access to the South African Rock Art Digital Archive, I would like to thank Azizo da Fonseca. His assistance in obtaining permission from numerous museums to allow us to publish their copyrighted images is also appreciated.

Special thanks is also given to Professor Karim Sadr for initiating this project between the Rock Art Research Institute and the Wits School of Computer Science.

I acknowledge the financial assistance from the Wits Centre of Excellence in Mathematical and Statistical Sciences for enabling me to attend and present a portion of this work at the SAICSIT 2014 conference in Centurion, Gauteng.

I also recognise that this research would not have been possible without the financial assistance of the National Research Foundation (NRF) under Grant 83692. All opinions expressed and conclusions arrived at are our own and should not necessarily be attributed to NRF.

Publications

Portions of this work have appeared in the publications listed below. For convenience, the documents can be seen in their entirety attached to the end of this document.

Title: Automatic Indexing of South African Rock Art
Authors: Amrit Purshotam, Turgay Celik, David Pearce, Azizo da Fonseca, Hima Vadapalli
Conference: Southern African Institute for Computer Scientists and Information Technologists (SAICSIT): Empowered by Technology, 28 Sept - 01 Oct 2014

Contents

Abstract i

Declaration ii

Acknowledgements iii

Publications iv

Contents v

List of Figures viii

List of Tables x

1 Introduction 1

2 Southern African Rock Art 3
- 2.1 Introduction . . . 3
- 2.2 Object breakdown . . . 3
  - 2.2.1 Geometric objects . . . 3
  - 2.2.2 Representational objects . . . 4
- 2.3 Artistic style breakdown . . . 6
  - 2.3.1 Engravings . . . 6
  - 2.3.2 Paintings . . . 6
  - 2.3.3 Sketches and re-paintings . . . 8
- 2.4 Intra-class variations . . . 9
- 2.5 Postures and viewpoints . . . 9
- 2.6 Partial occlusion . . . 9
- 2.7 Arbitrary rotations and scales . . . 12
- 2.8 Deterioration . . . 12
- 2.9 Conclusion . . . 14

3 Previous Work 15
- 3.1 Introduction . . . 15
- 3.2 Supervised learning . . . 15
- 3.3 Unsupervised learning . . . 15
- 3.4 Shape similarity descriptors . . . 17
- 3.5 Segmentation . . . 22

3.6 Visual analytics and crowdsourcing . . . 23
3.7 Applicability to this research . . . 23
3.8 Conclusion . . . 26

4 Object Detection Background **27**
4.1 Introduction . . . 27
4.2 Viola Jones object detection framework . . . 27
4.2.1 Haar-like features . . . 27
4.2.2 Integral image . . . 27
4.2.3 Tilted Haar-like features . . . 29
4.2.4 Selecting Haar features with AdaBoost . . . 30
4.2.5 Cascade architecture . . . 31
4.3 Support vector machines . . . 34
4.3.1 Linearly separable data . . . 34
4.3.2 Kernel trick . . . 37
4.3.3 Inseparable data . . . 38
4.4 Conclusion . . . 40

5 Research Methodology **41**
5.1 Introduction . . . 41
5.2 Aims . . . 41
5.3 Scope . . . 41
5.4 Programs and tools . . . 43
5.4.1 ImageClipper . . . 43
5.4.2 OpenCV . . . 43
5.4.3 LIBSVM . . . 43
5.5 Research questions . . . 43
5.6 Methodology . . . 44
5.6.1 Data collection and processing . . . 44
5.6.2 Parameter selection . . . 46
5.6.3 Experimental setup . . . 48
5.7 Conclusion . . . 48

6 Results **49**
6.1 Introduction . . . 49
6.2 Analysis of Viola Jones classifier . . . 49
6.2.1 True positive analysis . . . 49
6.2.2 False negative analysis . . . 51
6.2.3 False positive analysis . . . 52
6.2.4 True negative analysis . . . 54
6.3 Analysis of support vector machine classifier . . . 54
6.3.1 Performance on positive image set . . . 55
6.3.2 Performance on negative image set . . . 56
6.4 Analysis of classifiers on the SARADA . . . 56
6.4.1 Viola Jones performance . . . 57
6.4.2 Support vector machine performance . . . 59

6.4.3 Overall performance 59
6.5 Conclusion 60

7 Conclusion and Future Work **62**

Appendices **64**

A Viola Jones Cross Validation Results **64**

B Viola Jones False Positive Graphs **65**

C Viola Jones True Positives **67**

D Viola Jones False Negatives **71**

E Viola Jones False Positives **75**

F Viola Jones True Negatives **79**

G Support Vector Machine Cross Validation Results **83**

Bibliography **84**

SAICSIT 2014 Short Paper **91**

List of Figures

2.1 Geometric rock art . . . 4
2.2 Representational rock art . . . 5
2.3 Engraved rock art . . . 6
2.4 Artistic styles of painted rock art . . . 7
2.5 Historical sketches, redrawings, and paintings . . . 8
2.6 Intra-class variation of eland . . . 9
2.7 Postures and viewpoints of eland . . . 10
2.8 Postures of humans . . . 11
2.9 Partial occlusion of objects . . . 12
2.10 Arbitrary rotations and scales of objects . . . 13
2.11 Types of deterioration in the rock art . . . 14

3.1 Artificially constructed petroglyph scene . . . 16
3.2 Horn form petroglyph . . . 17
3.3 Petroglyph and its skeleton. . . . 18
3.4 Time-series with dynamic time warping and dendrograms of similar shapes 19
3.5 Pairs of similar petroglyphs . . . 20
3.6 Digitised and binarised petroglyph scene . . . 20
3.7 Skeletal graph creation process . . . 21
3.8 Result of late classifier fusion segmentation . . . 22
3.9 Edge segmentation filters . . . 24
3.10 Smoothing filters . . . 25

4.1 Rectangle Haar-like feature set . . . 28
4.2 Integral image . . . 28
4.3 Integral image application . . . 29
4.4 Rotated Haar-like feature set . . . 29
4.5 Adaboost algorithm for classifier learning . . . 31
4.6 Cascade architecture . . . 32
4.7 Cascade detector training algorithm . . . 33
4.8 Optimal hyperplane . . . 34
4.9 Transforming non-separable data . . . 37
4.10 Kernel trick . . . 39

5.1 Positive training set samples . . . 44
5.2 Negative training set samples . . . 45
5.3 Comparison of precision of 15 and 20 classifiers . . . 47

6.1 Viola Jones classifiers true positive rate summary . . . 50
6.2 Training set size vs. true positive rate . . . 50
6.3 Viola Jones classifiers false positives per image summary . . . 52
6.4 Positive training sample set size vs. number of false positives per image . 53
6.5 False positive breakdown of the Viola Jones classifiers . . . 54
6.6 Support vector machine performance on the positive test set . . . 55
6.7 Support vector machine performance on the negative test set . . . 56
6.8 Viola Jones eland classifier true positive rate and the number of false positives per image on the Simpson dataset compared to it's average performance in cross-validation . . . 57
6.9 False positive breakdown of Viola Jones classifier on the Simpson dataset 58
6.10 Viola Jones eland classifier precision on the Simpson dataset compared to its average performance in cross-validation . . . 58
6.11 Support vector machine performance on the Simpson dataset . . . 59
6.12 Breakdown of true positives on the Simpson dataset compared to the average during cross-validation . . . 60

B.1 Eland false positive analysis . . . 65
B.2 Elephant false positive analysis . . . 65
B.3 Humans false positive analysis . . . 66
B.4 Rhebuck false positive analysis . . . 66

C.1 Eland true positives . . . 67
C.2 Elephant true positives . . . 68
C.3 Human true positives . . . 69
C.4 Rhebuck true positives . . . 70

D.1 Eland false negatives . . . 71
D.2 Elephant false negatives . . . 72
D.3 Humans false negatives . . . 73
D.4 Rhebuck false negatives . . . 74

E.1 Eland false positives . . . 75
E.2 Elephant false positives . . . 76
E.3 Humans false positives . . . 77
E.4 Rhebuck false positives . . . 78

F.1 Eland true negatives . . . 79
F.2 Elephants true negatives . . . 80
F.3 Humans true negatives . . . 81
F.4 Rhebuck true negatives . . . 82

List of Tables

2.1 Representational object categories 4

5.1 Number of images and objects in each object class 42
5.2 Percentage of *flat* images 42
5.3 Object size ratios 45
5.4 Performance comparison of the Haar-like feature types 46
5.5 Performance comparison of different numbers of classifiers in the Viola Jones cascade 46
5.6 Performance comparison of the different boosting types 47
5.7 Average training and testing dataset sizes 48

A.1 Viola Jones cross-validation results 64

G.1 SVM cross-validation results 83

Chapter 1

Introduction

Rock art is the archaeological term used to describe pre-historic artworks placed on natural stone in rock shelters, caves, boulders, and platforms. It typically consists of engravings, paintings, drawings, stencils, and bas-relief carvings of abstract geometric figures and naturalistic representations of animals and humans. The recent discovery of 40 000 year old rock art in Indonesia [Aubert *et al.* 2014], comparable in age to the oldest European rock art, suggests that this form of human creativity emerged either independently across the world or had deeper origins thousands of years earlier from the first humans to leave Africa. Whatever the case, the art represents tens of thousands of years of ancient human experience, knowledge, and spirituality of generations of indigenous cultures around the world.

Closer to home, the Rock Art Research Institute (RARI) at the University of the Witwatersrand is spearheading efforts to understand the complex symbolism in southern African rock art. By linking what was known about San beliefs to the art, they found that it focuses on the experiences and sensations felt by San shamans and their actions through the spirit world while in trance. The capture of a rain animal (a mythological San creature) and the activation of forces for healing and fighting off enemies [Sarada 2014] are examples of such experiences. The RARI also found that the art itself held deep religious significance to the San. Eland blood was considered to be a powerful substance and so was mixed into the paints. As such, places with high concentrations of rock art were considered to be intensely powerful and spiritual to the San. These discoveries, and more, have led the RARI to become one of the largest rock art institutions in the world.

The RARI also specialises in the conservation and management of southern African rock art. They noticed that their extensive collection of historical documents and photographs were decaying, a problem further compounded by the deterioration of the original paintings themselves due to their exposure to natural forces such as wind and rain. As a major component of world history and human heritage, their disappearance would be a tremendous loss to future generations and humanity as a whole. The RARI, therefore, digitised their collections as well as the rock art sites themselves to enable further research after the art inevitably disappears forever. This has resulted in the South African Rock Art Digital Archive (SARADA), the world's largest rock art collection with over 275 000 images of paintings depicting a wide variety of objects such as humans and animals.

Each image in the SARADA has been tagged with meta-data indicating information such as the location of the original paintings and the institutions who currently hold

the copyright. However, information indicating the specific objects that appear in each image is not available. As a result, the RARI began the task of labelling each object in each image to enable text-based search queries on the SARADA. After labelling just 3%, it became evident that manual labelling was infeasible and that an automated classifier would be more suited to this task. A fully labelled SARADA would make this massive resource more useful to rock art researchers and thereby become a powerful tool that they could use to further progress in their field.

Therefore, in this work, we present the use of the Viola Jones object detection framework [Viola and Jones 2001] combined with a Support Vector Machine (SVM) [Cortes and Vapnik 1995] in a two-stage cascade to automatically classify and label the SARADA. The Viola Jones framework makes use of Haar-like features to encode knowledge about an object and, in constant time, decide whether to accept or reject a region of an image as the object. The SVM, on the other hand, makes use of a separating hyperplane to maximise the margin between classes of data (in our case, an image with or without the object). To test these techniques we implemented a five-fold cross-validation experimentation procedure on each of four common figures in the rock art namely eland, elephants, humans, and rhebuck. The positive datasets were created by manually cropping out the object from the original images while the negative datasets were created by randomly cropping out sub-windows that did not contain the object in it's corresponding positive dataset. To implement the cross-validation procedure, we then randomly split the positive and negative datasets into five distinct groups and trained a classifier for each group. Another groups positive and negative set was then used as the testing set for a given group however, the original source images were used instead of the cropped versions. The average performance for each of the four common figures was then calculated from the five classifiers and the results were found to be promising.

The rest of this dissertation is organised as follows. Chapter 2 describes the southern African rock art and its characteristics. Chapter 3 examines previous systems for segmentation, recognition, and analysis of rock art while Chapter 4 reviews the theory of the algorithms we selected for rock art recognition, the Viola Jones object detection framework and SVMs. Chapter 5 describes the research methodology while the analysis of the results is discussed in Chapter 6. To conclude, we answer the research questions and discuss potential future work in Chapter 7.

Chapter 2

Southern African Rock Art

2.1 Introduction

It is a common misconception that southern African rock art is a literal depiction of the daily lives of the San, the indigenous hunter-gatherer peoples of the region. Rock art is, in fact, considerably more complex and should be viewed as abstractions and representations of the visions and sensations felt by the San in a transcendent state known as trance. As an integral part of San culture, the rock art thus held deep religious and spiritual significance to them [Lewis-Williams 1990]. Due to the nature of this work, however, we will not delve deeper into the meanings and interpretations of the paintings but focus on the visual and cosmetic aspects of the art. Thus, this chapter will breakdown the various types of objects in the South African Rock Art Digital Archive in Section 2.2 and the different artistic styles of the paintings in Section 2.3. The characteristics of the objects, such as intra-class variations, different postures and viewpoints, partial occlusion, and arbitrary scales and rotations, are each discussed in Sections 2.4-2.7. Lastly, the reasons for the rapid deterioration of the paintings are explored in Section 2.8.

2.2 Object breakdown

At the top most level, there are two types of southern African rock art namely *geometric* and *representational* figures or objects.

2.2.1 Geometric objects

Geometric objects are characterised by abstract shapes and forms such as dots, crosses, zigzags, sets of lines, grids, vortices, etc. [Lewis-Williams 1990]. Examples of these objects can be seen in Figure 2.1. The origins of these shapes, as with most of San rock art, lie at the heart of the sacred trance dance, a ritual where shamans dance persistently while women clap and sing special medicine songs intensely. This ritual culminates in the shamans entering an altered state of consciousness known as trance. There are many stages of trance but in the first stage, these geometric forms are seen [Lewis-Williams *et al.* 1988].



(a) RARI-RSA-PLS1-22



(b) UOC-SRZ-01-3418P

Figure 2.1: Examples of geometric objects such as dots (2.1a) and wheel like structures (2.1b).
Images reproduced with permission from the Rock Art Research Institute, University of the Witwatersrand, South Africa (2.1a) and the University of Cologne, Germany (2.1b). Retrieved 13 October 2014, from SARADA, www.sarada.co.za.

2.2.2 Representational objects

Representational objects are recognisable as depictions of real life scenes or visions in deeper states of trance that appear to be real. They have been broadly classified in the SARADA as equipment, animals, humans, and the trance dance itself. The equipment class is composed of images of bows, arrows, quivers, bags, fly whisks, and digging sticks. The human class is composed of images of normal human figures while the trance dance class contains images of humans engaged in or experiencing trance. The animal class is the largest and contains images of several southern African wildlife. Examples of each class can be seen in Figure 2.2.

From Table 2.1, we can see that the number of images that have been classified into these representational object categories add up to 8111, a mere fraction (3% to be specific) of the 275 000 images in the SARADA. Of the 8111 images, we make use of the 5566 images in the human and animal categories.

Category	Number of Images
Equipment	1817
Animals	3685
Humans	1881
Trance Dance	715
Total	8111

Table 2.1: Number of images that have been classified into these representational object categories.



(a) RARI-RSA-WIE3-16D



(b) RARI-OLG-01-109



(c) RSA-HUT1-2



(d) IZI-LVF-01-406HC

Figure 2.2: Examples of representational objects in the SARADA of a human wielding a bow and arrow (2.2a), a lion from the animal class (2.2b), normal human figures (2.2c), and therianthropes, human-animal blends representing shamans in trance (2.2d).
Images reproduced with permission from the Rock Art Research Institute, University of the Witwatersrand, South Africa (2.2a, 2.2b, 2.2c) and the Iziko Museums of Cape Town, Social History Collections, South Africa (2.2d). Retrieved 14 October 2014, from SARADA, www.sarada.co.za.

2.3 Artistic style breakdown

2.3.1 Engravings

Engraved rock art is created by scraping, incising, and/or pecking at a rock surface instead of using any pigments. Examples of such engravings in the SARADA can be seen in Figure 2.3. Note that engraved rock art is sometimes referred to as *petroglyphs* and that we use these terms interchangeably in this document.



(a) RARI-ARW-01-1474H (b) RARI-RSA-UIT1-6

Figure 2.3: Examples of engraved rock art. In Figure 2.3a we can see two rhino, one of which is so carefully engraved, we can see the characteristic folds on its skin. Figure 2.3b shows a lion engraved by pecking at the rock.
Images reproduced with permission from the Rock Art Research Institute, University of the Witwatersrand, South Africa. Retrieved 13 October 2014, from SARADA, www.sarada.co.za.

2.3.2 Paintings

Painted rock art is created by applying pigments to a rock surface. The pigments are simply combinations of different natural substances that each produce a different colour. For example, red pigment is a mixture of ferric oxide and ochre, white from silica, china clay, and gypsum, and black from charcoal and specularite [Lewis-Williams 1990].

The number of colours and the way they are applied define four artistic styles, namely *monochrome*, *bichrome*, *polychrome*, and *shaded polychrome* paintings. Monochrome paintings consist of just one colour such as in Figure 2.4a where an eland is painted red. Bichrome paintings contain exactly two colours with a clear and rigid separation of colours. For example, Figure 2.4b depicts an eland painted in red and white. Polychrome paintings comprise of more than two colours, also with a clear distinction between colours such as in Figure 2.4c where an eland is depicted with an orange and white body with black horns, back stripe, tail-tuft, and hooves. Shaded polychrome paintings are also composed of more than two colours but the colours blend into each other. Figure 2.4d is an example of an eland where the orange, white, and black colours blend in around it's contours to create an almost photo-realistic illustration.



(a) RARI-RSA-INX1-80



(b) RARI-HPC-01-116HC



(c) RARI-HPC-01-44HC

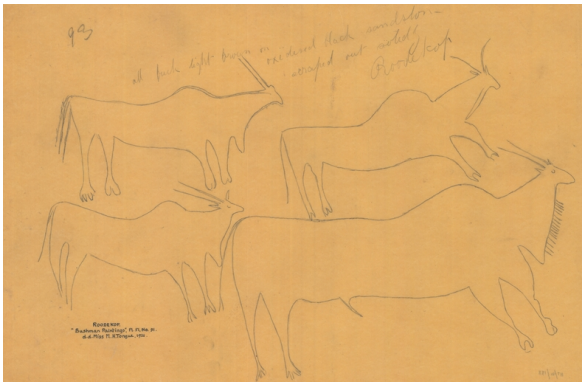


(d) RSA-BAE2-5

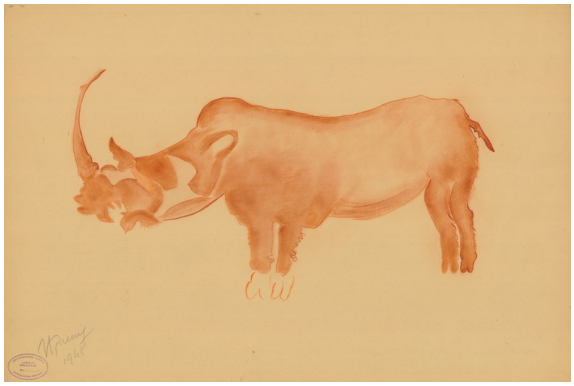
Figure 2.4: The four artistic styles of rock art paintings, monochrome (2.4a), bichrome (2.4b), polychrome (2.4c), and shaded polychrome (2.4d), are portrayed in these images of eland.
Images reproduced with permission from the Rock Art Research Institute, University of the Witwatersrand, South Africa. Retrieved 14 October 2014, from SARADA, www.sarada.co.za.

2.3.3 Sketches and re-paintings

The SARADA also contains copies of the numerous documents, sketches, and re-paintings from early 19th and 20th century colonial explorers and historians who visited the rock art sites. The sketches ranged from very basic outlines such as in Figure 2.5a to more detailed ones that were placed on drafting film such as in Figure 2.5c. Similarly, the paintings also ranged from the basic (2.5b) to the more intricate (2.5d).



(a) RARI-HT-01-188HC



(b) RARI-HB-01-346HC



(c) RSA-EBU1-1R



(d) NMSA-PJV-01-09HC

Figure 2.5: The various types of historical sketches, redrawings, and paintings in the SARADA. Figure 2.5a is a historical sketch of eland from 1914. Figure 2.5c is a redrawing of a therianthrope swarmed by bees originally placed on drafting film. Figure 2.5b is a basic oil painting of a rhino from 1948 while Figure 2.5d is a more intricate watercolour painting of an eland from 1956.
Images reproduced with permission from the Rock Art Research Institute, University of the Witwatersrand, South Africa (2.5a, 2.5b, 2.5c) and the Natal Museum (Department of Archaeology & Anthropology), Pietermaritzburg, South Africa (2.5d). Retrieved 15 October 2014, from SARADA, www.sarada.co.za.

2.4 Intra-class variations

The SARADA consists of rock art from across the southern African region. The paintings also range from 26 000 to just a few hundred years old. Thus, we can infer that many different cultures and generations of artists created the paintings. This has resulted in intra-class variations of an object with its features appearing in varying proportions from one to another. For example, the differences in appearance of an eland can be seen in Figure 2.6. The neck and torso of the eland in Figure 2.6a from the Mooi River is thick and well-built while the eland in Figure 2.6b from Lesotho is a lot skinnier.



(a) RARI-RSA-GAM1-417D



(b) LES-KKR5-3

Figure 2.6: These paintings from the Mooi River, KwaZulu-Natal (2.6a) and Lesotho (2.6b) illustrate the intra-class variation of eland with 2.6a thick and well-built and 2.6b a lot skinnier.
Images reproduced with permission from the Rock Art Research Institute, University of the Witwatersrand, South Africa (2.6a) and the Analysis of Rock Art in Lesotho Project, Lesotho (2.6b). Retrieved 15 October 2014, from SARADA, www.sarada.co.za.

2.5 Postures and viewpoints

The animals in the SARADA are positioned in a wide variety of postures and viewpoints. The most common is the side view but antelope, especially eland, are painted as seen from the front, back, and occasionally from above. In the side view especially, the animals are shown standing, running, leaping, lying down, and sometimes with their head turned [Lewis-Williams 1990]. Examples of these postures can be seen in Figure 2.7.

Human figures are similarly detailed and animated. They are depicted standing, walking, running, hunting, lying down, sitting, and dancing. Some are even drawn with their legs outstretched as if they are flying across the rock face [Lewis-Williams 1990]. Examples can be seen Figure 2.8.

2.6 Partial occlusion

Rock art painted directly over another painting is known as *super-positioning* in the rock art literature and results in the partial occlusion of the objects underneath. Ample



(a) LES-HAB1-84 (b) RARI-LEE-RSA-ARI1-5 (c) RARI-LEE-RSA-MIN1-2 (d) RARI-RSA-CHA1-55



(e) RARI-RSA-KAL2-1P (f) RARI-RSA-BUL1-45 (g) RARI-LEE-RSA-ROV3-29



(h) RSA-BAE2-88 (i) RSA-CLO1-28 (j) RARI-RSA-ELD1-114

Figure 2.7: The postures and views of eland vary from a top down view (2.7a, 2.7b), back view (2.7c, 2.7d), side profile views of them standing (2.7e), running (2.7f), leaping (2.7g), and laying down (2.7h), and with it's head turned (2.7i, 2.7j). However, by far the most common is the side profile standing view (2.7e).
Images reproduced with permission from the Analyis of Rock Art in Lesotho Project, Lesotho (2.7a) and the Rock Art Research Institute, University of the Witwatersrand, South Africa (2.7b-2.7j). Retrieved 15 October 2014, from SARADA, www.sarada.co.za.



(a) ZIM-LSR14-10 (b) JHH-01-583H (c) RARI-RSA-FRK1-36



(d) RARI-RSA-BUX1-27 (e) RARI-LEE-RSA-TYN2-1 (f) UKH-RSA-CSF31-3D

Figure 2.8: Differences between human postures in the rock art. They vary between standing (2.8a), walking (2.8b), running and leaping (2.8c), hunting (2.8d), sitting (2.8e), and dancing (2.8f).
Images reproduced with permission from the Rock Art Research Institute, University of the Witwatersrand, South Africa (2.8a, 2.8c, 2.8d, 2.8e), the John Hone Archive, KwaZulu-Natal, South Africa (2.8b), and the Tommy Topp Archive, South Africa (2.8f). Retrieved 16 October 2014, from SARADA, www.sarada.co.za.

free space on the rocks suggest that the artists intentionally painted over other art. Interestingly, statistical analyses indicate that certain combinations of super-positioning were favoured while others were avoided [Lewis-Williams 1990]. For example, as can be seen in Figure 2.9, eland were frequently painted on top of other eland or human figures but rarely the other way around. The reasons for this phenomenon are unknown. There are also cases in the art where paintings appear to be super-positioned but the object, supposedly underneath the first, was really painted in two parts.



(a) NMSA-PJV-01-173HC



(b) NMSA-PJV-01-39HC

Figure 2.9: The eland in Figure 2.9a and 2.9b are partially occluding the human figures behind them.
Images reproduced with permission from the Natal Museum (Department of Archaeology & Anthropology), Pietermaritzburg, South Africa. Retrieved 15 October 2014, from SARADA, www.sarada.co.za.

2.7 Arbitrary rotations and scales

The objects in the rock art are painted at arbitrary scales and rotations. The sizes may be an indication of the importance of an object while the rotations, especially if an animal is vertical or upside down, may represent a shaman that has completely transformed during trance. Examples can be seen in Figure 2.10.

2.8 Deterioration

The rock art is incredibly fragile and due to being present on rock caves, walls, and shelters, they are constantly exposed to climatic forces such as solar radiation, wind, and rain [Hœrlé and Salomon 2004; Hœrlé 2005 2006], geological forces such as the movement of water to and from the surface of rocks and of the rocks themselves [Van Rijssen 1987], and biological forces such as fungi, lichens, and mosses [Van Rijssen 1987]. Recent research has also shown that thermal variations in the pigments may cause pigment-to-pigment stresses within the painting itself [Hall *et al.* 2007]. All of these forces result in the paintings gradually and evenly fading, cracks appearing across them, and/or complete



(a) RARI-LEE-RSA-PEL5-17



(b) RARI-LEE-RSA-PEL5-50



(c) JHH-01-1734H



(d) RARI-ARW-01-847H

Figure 2.10: Figures 2.10a and 2.10b indicate upside down and vertical eland which may the sensation of a shaman that has completely transformed during trance. Figures 2.10c and 2.10d illustrates the arbitrary scales within and between images.
Images reproduced with permission from the John Hone Archive, KwaZulu-Natal, South Africa (2.10c) and the Rock Art Research Institute, University of the Witwatersrand, South Africa (2.10a, 2.10b, 2.10d). Retrieved 15 October 2014, from SARADA, www.sarada.co.za.

pieces of rock or paint flaking off. Examples of each type of deterioration can be seen in Figure 2.11.



(a) LES-HAB1-1443 (b) RSA-LAB11-10 (c) RARI-RSA-ANG3-2

Figure 2.11: The rock art appears to deteriorate by gradually and evenly fading away such as the leopard in 2.11a, flaking off of complete sections such as the torso of the human in 2.11b, and by cracks appearing across the art such as on the eland in 2.11c. Images reproduced with permission from the Analysis of Rock Art in Lesotho Project, Lesotho (2.11a) and the Rock Art Research Institute, University of the Witwatersrand, South Africa (2.11b, 2.11c). Retrieved 15 October 2014, from SARADA, www.sarada.co.za.

2.9 Conclusion

In this chapter, we reviewed various aspects of southern African rock art. We broke down the different types of objects and artistic styles in the art. We also examined the characteristics of the art and discovered that the same object can vary in appearance, posture, viewpoint, scale, rotation, be partially occluded by other objects. We also learned the reasons why the art is deteriorating and they way it manifests itself. In the next chapter, we will review the previous literature on the segmentation, recognition, and analysis of rock art.

Chapter 3

Previous Work

3.1 Introduction

Research into the recognition of rock art objects and symbols has been limited despite advances in image processing, machine learning, and computational power. To the best of our knowledge, the following chapter summarises all of the work carried out in this domain. We pay special attention to the algorithms and techniques used, the size and origin of the datasets, and the performance and efficiency of the algorithms employed. In Section 3.2, we discuss the lone supervised learning approach while Section 3.3 discusses the unsupervised learning approaches. In Section 3.4, we explore the various techniques that exploit shape similarity descriptors for the recognition and retrieval of similar petroglyphs. In Section 3.5, we detail an attempt at automatic segmentation of petroglyphs. In Section 3.6 we discuss visual analytics and crowdsourcing systems that support rock art researchers in the analysis and interpretation of rock art. We conclude this chapter by discussing the reasons these techniques and algorithms would not be applicable to our research in Section 3.7.

3.2 Supervised learning

A supervised learning algorithm analyses labelled training data to produce a function which can be used to classify new examples to one of the class labels.

There is only one such of example of this type of learning in relation to rock art recognition. Díaz and Castro [2002] studied the use of a neural network with back propagation to classify binarised images of Argentinian rock art. In their experiments they attempted to train their neural network to recognise three classes of patterns, however, their network suffered from oscillation with the last pattern trained always emerging as the output label when presented with new examples.

3.3 Unsupervised learning

An unsupervised learning algorithm analyses unlabelled training data in an attempt to find a hidden structure in the data and thereby create a classifier for new examples.

Deufemia *et al.* [2012a] used this idea and presented a Self-Organising Map (SOM) combined with a Fuzzy Visual Language Parser (FVLP) to classify petroglyphs from Mount Bego in the southern French Alps. An SOM is a type of artificial neural network that clusters data with an unsupervised competitive learning process. Clusters are organised in a two-dimensional topologically ordered feature map where common features are positioned geometrically close to each other. The FVLP, on the other hand, embeds basic knowledge about structures, patterns, and their relationships into inference rules which are grouped into a grammatical specification. The strategy thus employed was to use the SOM to produce a list of possible interpretations of a petroglyph while the FVLP exploited archaeological knowledge about recurring patterns within scenes of petroglyphs to refine that list. This was accomplished by training the SOM based on the *bag of visual words* framework [Dorkó and Schmid 2003] which consisted of three stages. First, the Radon transform [Rubin 2002] was selected as the shape descriptor of the petroglyphs and represented as numerical feature vectors. The SOM clustered the vectors with each cluster assigned an index value which was regarded as a word. Each symbol class was then represented by the frequencies of each visual word in its description. The FVLP encoded the archaeological knowledge about recurring patterns and groupings of certain petroglyphs within a visual language grammar. The evaluation of this strategy took place on a dataset of 74 binarised images from 11 classes of petroglyphs that were then rotated, translated, and skewed to generate a final dataset of 1406 images. The test set consisted of 50 scenes artificially constructed from combinations of 10 randomly selected petroglyph images, an example of which can be seen in Figure 3.1. When considering just the first item in the list of possible interpretations, 80.1% of the petroglyphs were correctly classified. When taking the top two interpretations, 90.7% of the petroglyphs were correctly classified.



Figure 3.1: Artificially constructed petroglyph scene of recurrent patterns that Deufemia *et al.* [2012a] attempted to classify.

Deufemia and Paolino [2013] later tried a combination of an SOM with an Image Deformation Model (IDM) where, as before, the SOM produced a quick first classification and the slower IDM refined it. This time, the SOM clusters were produced using a shape context descriptor [Belongie *et al.* 2002]. The method identifies n points on the shape contour and connects each point to all other points through $n - 1$ vectors with the relative distances measured in a uniform log polar space. The IDM [Keysers *et al.* 2004], meanwhile, computes the distances of each pixel of a query image with the corresponding pixels of the target image located within a certain warp range by taking into account the surrounding pixels. The evaluation took place on a dataset of 51 binarised images belonging to 17 classes that were translated, rotated, and skewed to produce a final dataset of 1530 images, an example of which can be seen in Figure 3.2. The SOM and IDM combination achieved an accuracy of 68.6% which took 141 minutes to compute.

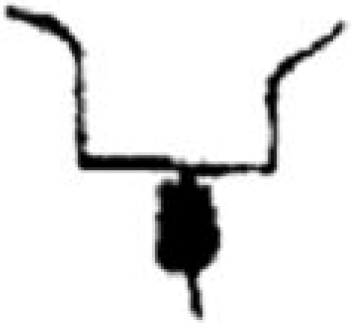


Figure 3.2: An example of a digitised and binarised petroglyph representing a horn form from Mount Bego that Deufemia and Paolino [2013] attempted to classify.

3.4 Shape similarity descriptors

A shape similarity descriptor is an alternative representation of a shape or image that allows the recognition and retrieval of visually similar shapes that may not necessarily be mathematically identical.

Takaki *et al.* [2006] developed a shape similarity descriptor for each type of petroglyph and analysed the statistics of groups of central Asian petroglyphs, such as the number of animals and humans on a single rock. We will only focus on their shape similarity descriptor as it is more relevant to our research. To characterise the shape of the petroglyph the authors first processed the binary image to produce a skeleton, an example of which can be seen Figure 3.3. Each body part of the skeleton was then expressed by alphabetic letters to denote, in the case of ibexes, the nose (N), horns (H), legs (L), body (B), and tail (T). Each petroglyph of the given object was then encoded by this scheme and the differences in this string were noted to produce a similarity measure. The authors experimented on ibex petroglyphs and thought their results to be "promising".

Wei *et al.* [2007] present rotationally invariant shape matching algorithms for classifying and clustering anthropological objects such as skulls, projectile points, and petroglyphs from Nevada and California. They first convert their shapes to a time-series, a one-dimensional shape representation of a two-dimensional shape where the distance from every point on the profile to the centre is measured and plotted on the y-axis. Dynamic

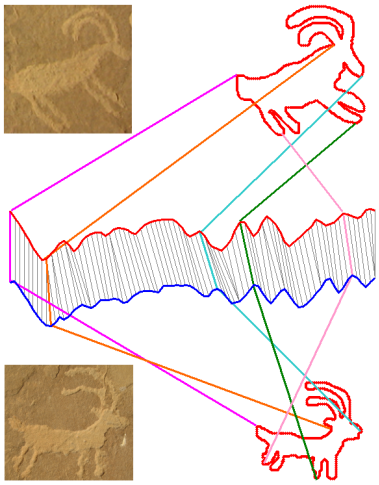


(a) Petroglyph relief (b) Skeleton

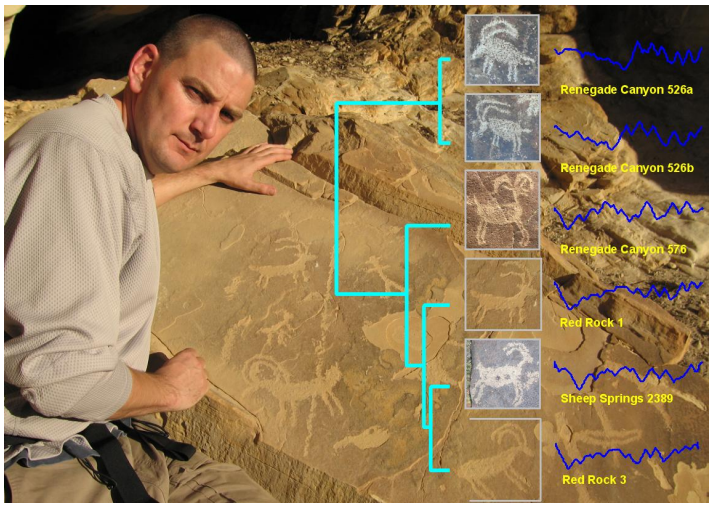
Figure 3.3: Example of a petroglyph and its skeleton [Takaki *et al.* 2006].

Time Warping (DTW) [Keogh and Ratanamahatana 2005; Ratanamahatana and Keogh 2005] was then used to measure the distance between two time-series since it allows the alignment of two sequences to stray from each other to compensate for local distortions in each shape. An example of DTW can be seen in Figure 3.4a. Since the paper focused on speeding up the calculations of distances measures, the experimentation on the accuracy of these techniques is limited and the reader is referred to previous literature [Gdalyahu and Weinshall 1999; Adamek and O'Connor 2003 2004; Attalla and Siy 2005; Jalba *et al.* 2005; Ratanamahatana and Keogh 2005; Vlachos *et al.* 2005]. On the non-petroglyph datasets, an average error rate of 13.58% was achieved. With petroglyphs, the authors simply remark that on "large" collections they achieve very "intuitive" results as in the dendrograms in Figures 3.4b and 3.4c, where similar shapes are clustered together. However, the image processing software for automatically extracting the petroglyph profiles failed on weather degraded ones resulting in unintuitive results.

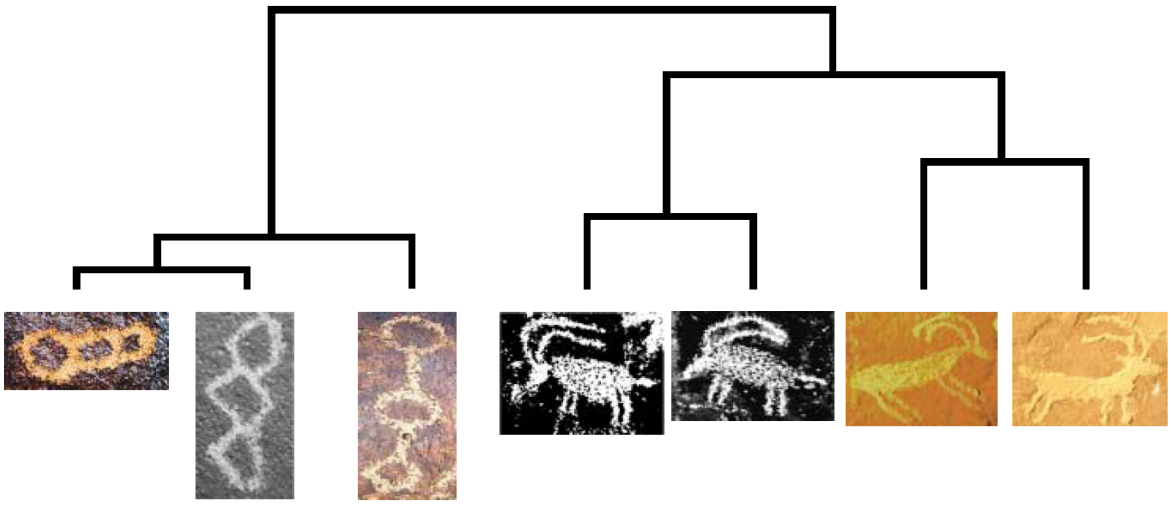
Perhaps the most in depth study of measuring similarity in petroglyphs was carried out by Zhu *et al.* [2008 2009 2011]. The similarity descriptor they presented is a variation of a Generalized Hough Transform (GHT). The algorithm took in a down-sampled binary outline/profile of a petroglyph and efficiently calculated the minimum number of pixels that did not superimpose themselves over the profile of a target petroglyph. Three experiments were performed to test the utility, accuracy, and scalability of this algorithm. The utility was tested on a set of 2852 southern Californian petroglyphs that were scanned from the book *Indian Rock Art of southern California* by Smith and Turner [1975] and then binarised, thinned, and down-sampled. Out of a possible 4 065 526 pairs, the GHT found 52 pairs within 18 minutes, five of which can be seen in Figure 3.5. The accuracy was tested on four large publicly available datasets that resemble petroglyphs, the NicIcon dataset (24 441 images split into 14 categories) [Niels *et al.* 2008], Farsi digits (102 352 images split into 10 categories) [Khosravi and Kabir 2007], music symbols of clefs (2128 images split into 4 categories) [Fornés 2009], and architectural symbols (7414 images split into 50 categories) [Sánchez *et al.* 2004]. On average, the accuracy of the GHT on these sets was 93.35%. To test scalability, a synthetic dataset was created by selecting 22 petroglyphs which were duplicated by 10 volunteers drawing on tablets to create a total



(a) Dynamic time warping



(b) Dendrogram of clustered ibexes.



(c) Similar geometric figures and ibexes are clustered into a dendrogram [Wei *et al.* 2007].

Figure 3.4: Figure 3.4a illustrates the time-series of two ibexes and the result of dynamic time warping where non-aligned points are matched. Figures 3.4b and 3.4c show the best-matches clustered into dendrograms with Figure 3.4b also showing the shapes' time-series to the right.
Figures 3.4a and 3.4b © Eamonn Keogh (2007), *Shape Matching*. Retrieved 12 November 2014, from University of California Riverside, www.cs.ucr.edu

of 250 petroglyphs. A second order polynomial transformation was then applied to each petroglyph to create 8 datasets of 10 000 to 1 280 000 images. The GHT was found to be at least 712 times faster than brute force search as 99.84% of the calculations are pruned.

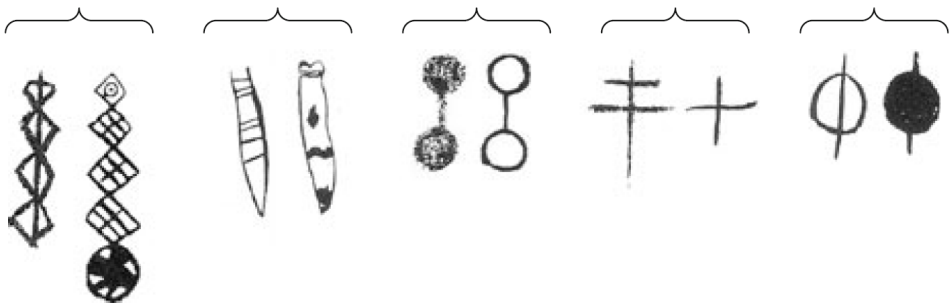


Figure 3.5: Five pairs of similar petroglyphs discovered by the Generalized Hough Transform [Zhu *et al.* 2011].

Deufemia and Paolino [2014] experimented with another shape similarity measure, the Enhanced Generic Fourier Descriptor (EGFD), which is invariant to translation, rotation, scaling, and distortions. To allow this, the images were first normalised based on a procedure by Zhang and Lu [2002a]. The EGFD's were subsequently extracted by applying a modified polar Fourier transform on a rectangular polar shape image [Zhang and Lu 2002b] and normalised based on their radial and angular frequencies. Similarity between two shapes was then measured with the Euclidean distance between their feature vectors against a known dataset of 94 images from 10 classes. The evaluation took place on a dataset of 53 scenes containing 22 petroglyphs from Mount Bego, an example of which can be seen in Figure 3.6. Almost half the petroglyphs, or 46%, were unrecognisable and from the other 54% of the petroglyphs, 65% were recognised correctly.



Figure 3.6: Digitised and binarised scene from Mount Bego where Deufemia and Paolino [2014] segmented each petroglyph and subsequently attempted to classify them.

Seidl *et al.* [2014] proposed to model petroglyphs as skeletal graphs [Bai *et al.* 2007] in order to reduce the petroglyph similarity problem to a graph similarity one. The algorithm used each end point and each junction point of the resulting graph as nodes and the skeleton branches connecting these points as edges with short edges pruned. An example of this process can be seen in Figure 3.7. Graph Embedding (GE) is subsequently used to create a feature vector for each graph by extracting 21 graph properties, such as the number of nodes, edges, cycles, etc., and the pairwise distances between graphs are calculated with Euclidean distances. Classification is then performed using the k-nearest neighbours algorithm. The evaluation of this approach took place on a dataset of 100 digitised and binarised tracings of petroglyphs belonging to 10 classes from the Valcamonica site. Using leave one out cross-validation and creating a classifier for each of the 21 graph properties, they found that the classifier using the *graph energy* property (which is the sum of the absolute values of the real components of the eigenvalues) had a classification accuracy of 47%. However, fusing 4 and 5 of the best performing properties as another feature vector produced a classification accuracy of 57%. They then took the fusion strategy further and fused the GE descriptor with the Inner Distance Shape Context (IDSC) [Ling and Jacobs 2007], Shape Context [Belongie *et al.* 2002], and the GHT [Zhu *et al.* 2011] descriptors to produce an 88% classification accuracy.

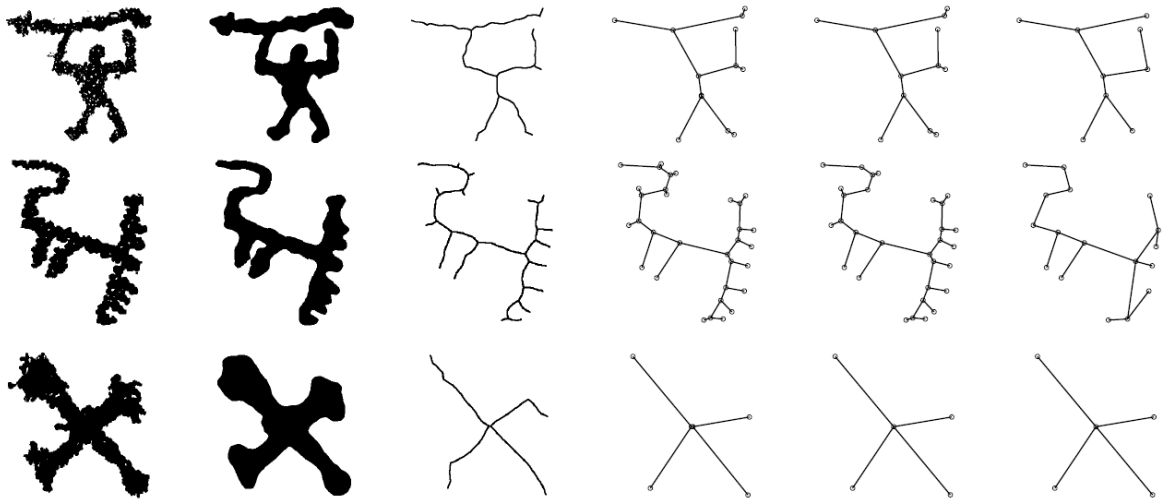


Figure 3.7: Examples of the process to create the skeletal graphs used by [Seidl *et al.* 2014]. The figure shows the petroglyph tracings, the pre-processed images, the skeleton, and then the skeletal graphs with different pruning thresholds of lines with less than 2, 10, and 30 pixels respectively.

Mascardi *et al.* [2014] developed the Indiana Multi-Agent System composed of three components, AgentRec, ImageRec, and Manent for sketch recognition, image recognition, and multi-lingual text processing respectively. The component related to our research is the ImageRec component where the authors propose the use of the Viola and Jones [2001] object detection framework and SIFT descriptors for recognition of colour petroglyphs but relegate their investigation to future work since the blurriness, sketchiness, and incompleteness of the petroglyphs made the task too difficult for them. They instead focused on the easier task of recognising digitised and binarised petroglyph reliefs from the Mount

Bego site. They investigated the use of the Radon transform [Rubin 2002] and the Shape Context [Belongie *et al.* 2002] as their shape descriptors and train separate SOM's with each descriptor on each of 11 classes of petroglyphs. The shape context descriptor was found to perform best with the *reticulate* and *zigzag* classes with a classification accuracy of 83% while it performed worst on the *oxen* class with an accuracy of 17%. Meanwhile, the Radon transform performed the best on the *personage* class with an accuracy of 78% and worst on the *goddess* class with an accuracy 51%.

3.5 Segmentation

Seidl and Breiteneder [2011 2012] presented a support vector machine classifier fusion strategy for segmenting petroglyphs, from Capo di Ponti, Valcamonica, by pixel classification. They argue that the most important features to enable segmentation of petroglyphs are colour, edges, and textures. Thus, they trained three separate support vector machines using SIFT [Lowe 2004], gray level co-occurence matrices [Haralick *et al.* 1973], and Local Binary Patterns [Ojala *et al.* 2002] respectively and tested the performance of each. They then tested combinations by concatenation of the above features and measured the performance. They also experimented with combinations of the output labels of the best performing classifiers from above and employ a majority voting procedure with these classifiers. The last strategy, or *late fusion* as they call it, was found to be the best and its performance ranged from a 96% recall with 42% precision to 41% recall with 88% precision. An example of a successful segmentation with the late fusion classifier can be seen in Figure 3.8.

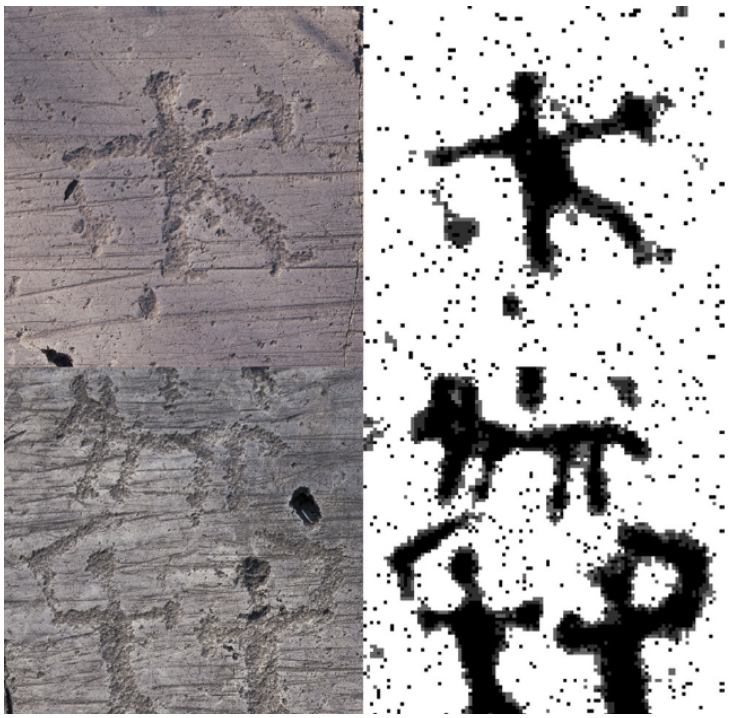


Figure 3.8: Result of late classifier fusion with thresholded majority voting. Darker pixels correspond to the confidence level [Seidl and Breiteneder 2012].

3.6 Visual analytics and crowdsourcing

Visual analytics is an outgrowth of the fields of information visualisation and scientific visualisation that focuses on analytical reasoning facilitated by interactive visual interfaces [Thomas and Wong 2004] while crowdsourcing is the process of obtaining services and content from a large group of volunteers rather than employees or suppliers. Since these are not classification systems, this section will be kept brief and is only included for completeness.

Deufemia *et al.* [2012b] presented Indiana Finder, an interactive visualisation system that assists archaeologists in the examination of large repositories of information and drawings by providing visual analytic support for the interpretation of new archaeological findings, detection of interpretation anomalies, and the discovery of new insights.

Zhang *et al.* [2013] developed a visual analytics framework and a set of tools for the discovery of degradation patterns on wall paintings.

Deufemia *et al.* [2014b] put forward another visual analytics system, named DARK (Discovery of Ancient Rock art Knowledge), that combines visualization techniques with fuzzy-based analysis of rock art scenes to infer information for the correct interpretation of scenes to allow archaeologists to validate their hypothesis against information stored in the repository.

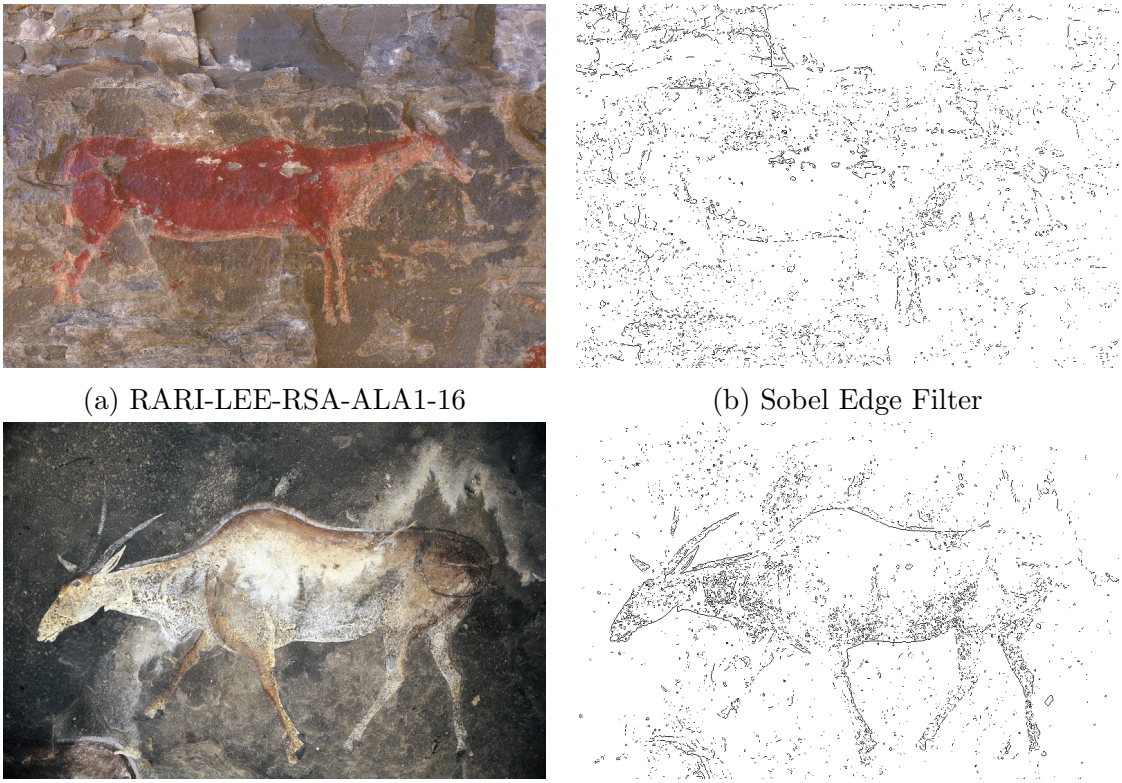
In the second part of the research by [Zhu and Keogh 2010; Zhu *et al.* 2011], they present CAPTCHA-ROCK, a system that outsources the difficult petroglyph feature extraction task to humans. They frame the problem as a CAPTCHA (Completely Automated Public Turing Test to tell Computers and Humans Apart) [Von Ahn *et al.* 2003] where volunteers manually trace the contours of each petroglyph which can then be passed on to their GHT algorithm for classification.

Deufemia *et al.* [2014a] further expanded on this concept and present PetroAdvisor, a mobile application supporting several fundamental activities to digitally preserve petroglyph sites. The system exploits a rewarding strategy to stimulate participation so that people who enter useful information such as petroglyph pictures, descriptions, and contours can then access other archaeological data, tips on excursions and tours, and ratings from previous tourists.

3.7 Applicability to this research

The techniques and algorithms described above are only applicable to well defined images of abstracted and symbolic-like petroglyphs that have been segmented from the rock surfaces and then binarised to produce clean images with no background noise. The only exception here would be the work by Seidl and Breiteneder [2011 2012]. Some of the datasets were further limited to images containing just a single petroglyph. The SARADA dataset, on the other hand, consisted of representational paintings of animals and humans. Most importantly, the images in the SARADA are the raw photographs of these rock paintings which include all the background noise of variably coloured and jagged rock surfaces. The paintings themselves were also not always well defined as they were faded, flaked, or cracked. The images also contained multiple paintings of different objects at arbitrary scales and rotations unlike most of the datasets used above which limit an image to just one object.

In an attempt to simplify and clean up the images (like the ones in the previous literature) we applied standard image processing algorithms for segmentation and smoothing. However, they did not perform well when applied to our rock art datasets as they required fine tuning of their parameters to produce an optimal result. To demonstrate this we attempted to segment out an eland from two images using the Sobel [Sobel 1990], Prewitt [Prewitt 1970], Roberts [Roberts 1963], Laplacian of Gaussian [Sotak and Boyer 1989], and Canny [Canny 1986] methods. The Sobel method performed the best but as illustrated in Figure 3.9 this method (and the others) did not generalise well to all rock art images. The eland in Figure 3.9d was segmented reasonably well although its front leg and rump are missing while in the eland in Figure 3.9b, no discernible edges can be seen. Both images also exhibit background noise from the rock surface.



(a) RARI-LEE-RSA-ALA1-16

(b) Sobel Edge Filter

(c) RSA-BAE2-6

(d) Sobel Edge Filter

Figure 3.9: The Sobel edge segmentation filter is not generalisable when applied to rock art. In a low contrast image (in grayscale) (3.9a), no discernible edges of the eland were extracted (3.9b) while in a higher contrast image (3.9c), edges do seem to have been extracted but also noise from the background cave wall (3.9d).
Images reproduced with permission from the Rock Art Research Institute, University of the Witwatersrand, South Africa. Retrieved 25 October 2014, from SARADA, www.sarada.co.za.

We also applied two smoothing filters to the image in Figure 3.10a, the standard Gaussian smoothing and the edge preserving anisotropic diffusion [Perona and Malik 1990], in an attempt to simplify the features in the image and remove as much background noise as possible. The result of the Gaussian filter can be seen in Figure 3.10b and we can see

that it blurred the image but the horse did not blend into a solid red colour nor did the background fuse into a greyish-brown colour. On the other hand, the result of the anisotropic diffusion filter can be seen in Figure 3.10c and we can see that it did manage to accomplish this to some degree although the spots of rock on the horse do seem to have been emphasized. Again these methods do not generalise well to all rock art images since their parameters have to be fine tuned for each image to produce acceptable results.



(a) NMSA-JHH-01-277H (b) Gaussian filter (c) Anisotropic Diffusion

Figure 3.10: An image of a horse (3.10a) was smoothed over by the standard Gaussian filter (3.10b) and the edge preserving anisotropic diffusion filter (3.10c).
Images reproduced with permission from the Natal Museum (Department of Archaeology & Anthropology), Pietermaritzburg, South Africa. Retrieved 26 October 2014, from SARADA, www.sarada.co.za.

These differences in the dataset and the inability to simplify them required a different strategy to overcome these unique obstacles. We considered the use of the Scale Invariant Feature Transform (SIFT) [Lowe 1999 2004] where key points of an image are extracted from a set of sample images of an object which are then compared to key points of a new image based on the Euclidean distance of their feature vectors. However we found that this technique would not be suitable since it's strength was at matching the same object at different transformations. For similar reasons we also discounted the use of Speeded Up Robust Features (SURF) [Bay *et al.* 2006], a technique which uses an integer approximation of the determinant of Hessian blob detector. Another technique we found promising, by Bagon *et al.* [2010] was the use of a self-similarity descriptor to generate a model of an object which could then be used to find matches in new images. However, in preliminary testing of this technique we found that this technique required the user to locate the objects for it before the model could be generated or a match to take place.

We eventually selected the Viola and Jones [2001] object detection framework to provide a quick first classification through the use of Haar-like features, which through the use of the integral image, allowed them to be calculated in constant time. Speed was an important factor considering that there are hundreds of thousands of images to be classified. The Haar-like features encode knowledge about an object in difficult domains from a finite set of training data with the resulting classifier invariant to scale, can detect multiple instances of an object in a single image, generalises well to intra-class variations of an object, and can be trained to reject rocky background noise.

Inevitably, due to the difficulty and number of obstacles the Viola Jones classifier

attempted to overcome, mistakes did slip through. Therefore, we employed a Support Vector Machine [Cortes and Vapnik 1995] that provided a secondary classification of the sub-windows detected by the Viola Jones classifier as it had the simpler task of looking for a single object at a fixed scale.

3.8 Conclusion

In this chapter, we learnt about the various approaches that have been applied to the field of rock art object detection and clustering. The supervised and unsupervised learning methods were described as well as efforts to exploit shape similarity descriptors for recognition and clustering. Furthermore, a method for the automatic segmentation of petroglyphs was described. For completeness, we also briefly outlined the visual analytics and crowdsourcing systems that aid archaeological researchers in the analysis and interpretation of rock art. However, we also learnt that that most of the above systems made use of simplified datasets of manually segmented and binarised petroglyphs, unlike the SARADA dataset which is composed of raw photographs of colour paintings. This difference necessitated the need for a different approach to detect and recognise the southern African rock art objects namely the combination of the Viola Jones object detection framework and a support vector machine classifier. In the next chapter, we outline the theory of these algorithms and describe how they work.

Chapter 4

Object Detection Background

4.1 Introduction

In this chapter, we focus on the computer vision and pattern recognition methods of this research. We describe the Viola Jones object detection framework which makes use of Haar-like features to encode knowledge about an object and decide whether to accept or reject a region in an image as the object in Section 4.2. We also derive the formulation of the Support Vector Machine which makes use of a separating hyperplane to maximise the margin between classes of data in Section 4.3.

4.2 Viola Jones object detection framework

4.2.1 Haar-like features

Haar-like features are simple digital image features. Oren *et al.* [1997] and Papageorgiou *et al.* [1998] were the first to use a feature set based on Haar wavelets [Haar 1910], instead of image intensities, for pedestrian and face detection respectively. Viola and Jones [2001] then adapted this idea and developed Haar-like features. Here, the pixel intensities of adjacent rectangular regions at a specific location are summed up and subtracted from each other. This difference is then used to classify subsections of an image. Examples of such rectangular features are given in Figure 4.1.

4.2.2 Integral image

The integral image [Crow 1984] is an intermediate representation of an image that allows the sum of rectangular features to be computed in constant time [Viola and Jones 2001]. The integral image at location (x, y) contains the sum of pixels above and to the left of (x, y) inclusive:

$$I(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y')$$

where $I(x, y)$ is the integral image and $i(x, y)$ is the original image at (x, y). A visualisation of this concept can be seen in Figure 4.2. Moreover, the integral image can be

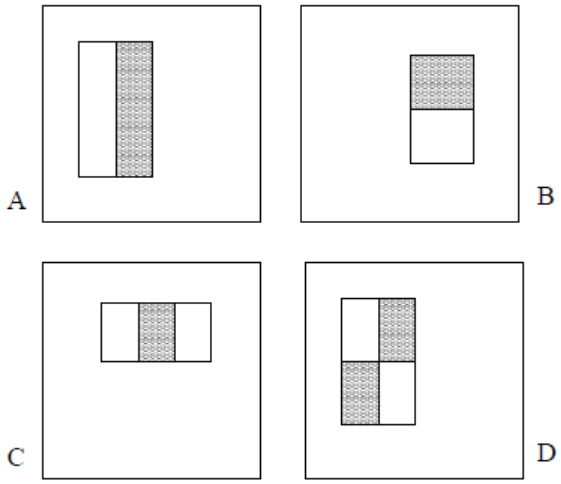


Figure 4.1: The sum of the pixels which lie within the white rectangles are subtracted from the sum of pixels in the grey rectangles. Two-rectangle features are shown in (A) and (B). (C) and (D) show three and four-rectangle features respectively [Viola and Jones 2001].

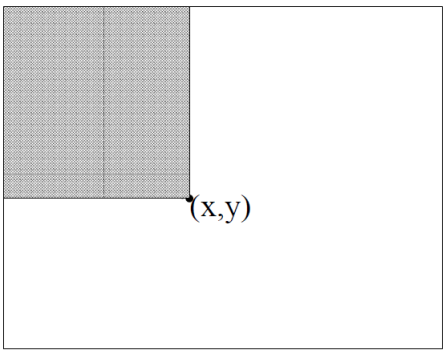


Figure 4.2: The value of the integral image at point (x, y) is the sum of all the pixels above and to the left [Viola and Jones 2001].

computed in a single pass over the original image using the fact that the integral image at (x, y) is just:

$$I(x, y) = i(x, y) + I(x - 1, y) + I(x, y - 1) - I(x - 1, y - 1)$$

Any rectangular sum can be computed in four array references regardless of its position or scale. The difference between two rectangular sums can be computed in eight array references. However, since the two rectangle features defined above involve adjacent rectangular sums, they can be computed in six array references, eight in the case of three-rectangle features, and nine for four-rectangle features as can be seen in Figure 4.3 [Viola and Jones 2001].

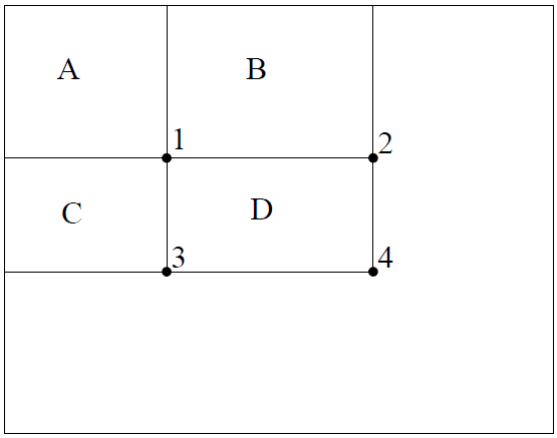


Figure 4.3: The sum of the pixels within rectangle D can be computed with four array references. The value of the rectangle image at location 1 is the sum of the pixels in A. The value at 2 is $A + B$, at 3 is $A + C$, and at 4 is $A + B + C + D$. The sum within D can be computed as $4 + 1 - (2 + 3)$ [Viola and Jones 2001].

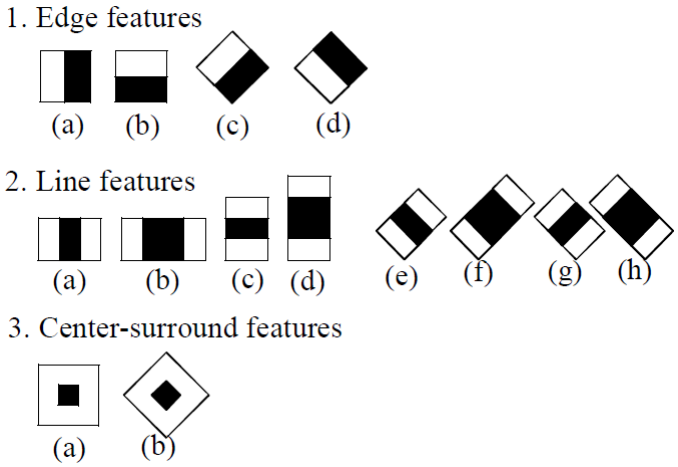


Figure 4.4: The 14 Haar-like features used by Lienhart *et al.* [2003].

4.2.3 Tilted Haar-like features

The rectangular Haar-like feature set used by Viola and Jones [2001] are all oriented either horizontally or vertically. These features work well to encode vertical and horizontal lines or edges in an image but not ones that are tilted. Thus, Lienhart *et al.* [2003] extended the original Haar-like feature set by including a set of 45° rotated and center surround features which improved the performance of the object detection system. Importantly, these new features can also be computed at all scales in constant time.

As seen in Figure 4.4, there are now 14 feature prototypes:

- four edge features
- eight line features
- two center-surround features

(feature D in Figure 4.1 has been discarded since it is well approximated by features (e) and (g) in Figure 4.4). These prototypes are scaled independently in the vertical and horizontal directions in order to generate a rich, over complete set of features (note that the line features can be calculated by two rectangles only).

To calculate these features at any scale in constant time, two intermediate representations of the image are required. The first is identical to the integral image calculation as described in Section 4.2.2 for all the upright rectangle features. For the 45° rectangles, the intermediate representation is the *Rotated Integral Image*. It is defined as the sum of the pixels of a 45° rotated rectangle with the bottom most corner at (x, y) and extending upwards until the boundaries of the image:

$$RI(x, y) = \sum_{y' \leq y, y' - |x - x'|} i(x', y')$$

It can also be calculated in one pass from left to right and top to bottom over all pixels by:

$$\begin{aligned} RI(x, y) = {} & RI(x - 1, y - 1) + RI(x + 1, y - 1) \\ & - RI(x, y - 2) + i(x, y) + i(x, y - 1) \end{aligned}$$

with $RI(-1, y) = RI(x, -1) = RI(x, -2) = RI(-1, -1) = RI(-1, -2) = 0$. From this the pixel sum of any rotated rectangle $r = (x, y, w, h, 45°)$ can be determined by four table lookups [Lienhart *et al.* 2003]:

$$\begin{aligned} RecSum = {} & RI(x - h + w, y + w + h - 1) + RI(x, y - 1) \\ & - RI(x - h, y + h - 1) - RI(x + w, y + w - 1) \end{aligned}$$

4.2.4 Selecting Haar features with AdaBoost

Adaptive Boosting, or AdaBoost for short, is a machine learning algorithm introduced by Freund and Schapire [1995] to improve the performance of other learning algorithms. AdaBoost builds a sequence of classifiers where the ith classifier is biased to correct the misclassification of the $1, \ldots, i - 1$ classifiers. Thus, the training error of the strong classifier approaches zero exponentially [Freund and Schapire 1995].

Viola and Jones [2001] use a variant of Adaboost to both select a small set of features that combine to form an effective classifier and to train the classifier itself. As shown in Figure 4.5, a weak learning algorithm selects the single rectangle feature which best separates the positive and negative examples. For each feature, the weak learner determines the optimal threshold classification function, such that the minimum number of examples are misclassified. A weak classifier $h_j(x)$ thus consists of a feature f_j, a threshold θ_j, and a parity p_j indicating the inequality sign:

$$h_j(x) = \begin{cases} 1 & \text{if } p_j f_j(x) < p_j \theta_j \\ 0 & \text{otherwise} \end{cases}$$

where x is a sub-window of an image.

- Given example images $(x_1, y_1), \ldots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive examples respectively.
- Initialise weights $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negatives and positives respectively.
- For $t = 1, \ldots, T$
  1. Normalise the weights
  $$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}} \tag{4.2.4.1}$$
  so that w_t is a probability distribution
  2. For each feature, j, train a classifier h_j which is restricted to using a single feature. The error is evaluated with respect to w_t, $\epsilon_j = \sum_i w_i |h_j(x_i) - y_i|$.
  3. Choose the classifier, h_t, with the lowest error ϵ_t.
  4. Update the weights
  $$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i} \tag{4.2.4.2}$$
  where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$
- The final strong classifier is:
$$h_j(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise} \end{cases} \tag{4.2.4.3}$$

where $\alpha_t = log \frac{1}{\beta_t}$

Figure 4.5: The AdaBoost algorithm for classifier learning. Each round of boosting selects one feature from all the potential features [Viola and Jones 2001].

4.2.5 Cascade architecture

Viola and Jones [2001] noticed that within any single image, the majority of the sub-windows are negative. A simple and efficient classifier is used to reject most of the negative sub-windows while letting the positive ones trigger the second, slightly more complex classifier. A positive sub-window from this classifier will trigger the third, even more complex classifier and so on. Thus, a positive instance triggers every classifier in the *cascade* while a negative sub-window is immediately rejected where it is detected. Figure 4.6 illustrates this concept.

Training this cascade of classifiers involves a number of trade-offs. In particular, classifiers with more features will achieve higher detection rates and lower false positive rates but at the same time require more time to compute. Thus, the number of classifiers, the number of features in each classifier, and the thresholds of each stage, are traded off in order to minimise the number of evaluated features.

In order to achieve this, a minimum acceptable detection rate and a maximum acceptable false positive rate per layer are selected. For example, a minimum detection rate

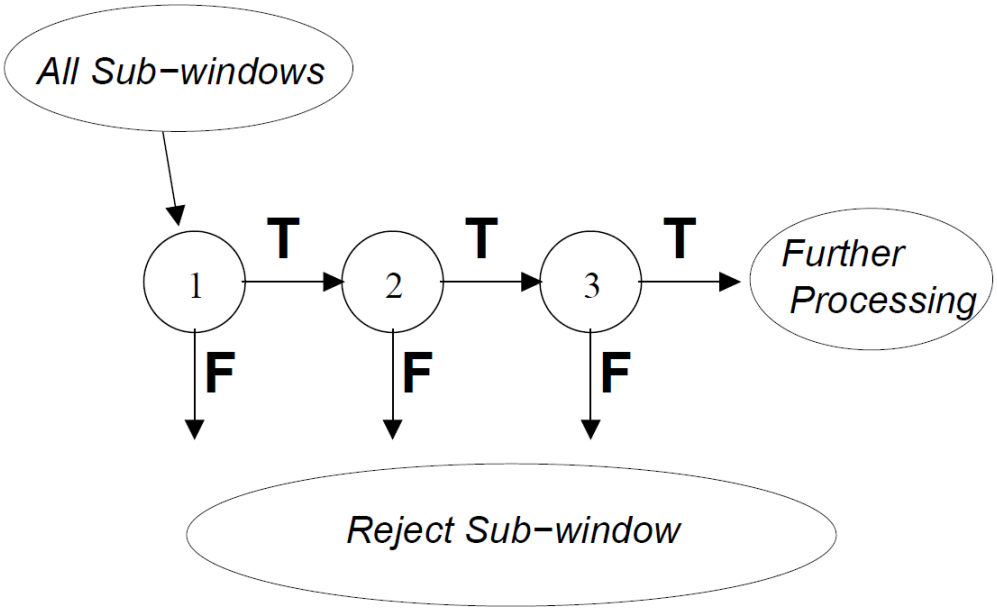


Figure 4.6: A series of successively more complex classifiers are applied to every sub-window. Negative sub-windows are eliminated as soon as they are discovered. *Further processing* can be additional classifiers or an alternate detection system [Viola and Jones 2001].

of 99% and a maximum false positive rate of 30% per layer for a 10 layer cascade ensures an overall detection rate of $0.99^{10} \approx 90\%$ and a false positive rate of $0.3^{10} \approx 10^{-6}$. Each layer of the cascade is trained by AdaBoost with the number of features increased until the target detection and false positive rates are met. If they are not met, another layer is added to the cascade. Figure 4.7 describes this algorithm in detail.

- User selects for f, the maximum acceptable false positive rate per layer and d, the minimum acceptable detection rate per layer
- User selects target false positive rate, f_{target}
- P = set of positive examples
- N = set of negative examples
- $F_1 = 1, D_0 = 1$
- $i = 0$
- while $F_i > F_{target}$
  - $i \leftarrow i + 1$
  - $n_i = 0, F_i = F_{i-1}$
  - while $F_i > f \times F_{i-1}$
    - $n_i \leftarrow n_i + 1$
    - Use P and N to train a classifier with n_i features using AdaBoost
    - Evaluate current cascaded classifier on validation set to determine F_i and D_i
    - Decrease threshold for the ith classifier until the current cascaded classifier has a detection rate of at least $d \times D_{i-1}$ (this also affects F_i)
    - $N \leftarrow \emptyset$
    - If $F_i > F_{target}$ then evaluate the current cascaded detector on the set of non-object images and put any false detections into the set N

Figure 4.7: The training algorithm for building a cascaded detector [Viola and Jones 2001].

4.3 Support vector machines

Support Vector Machines (SVMs) are supervised learning models developed by Cortes and Vapnik [1995] that analyse data and recognise patterns for data classification.

The derivation of the generalised support vector machine is split into three parts and is based on the tutorials by Burges [1998] and Kroon and Omlin [2004]. The base case of training an SVM on linearly separable data is explained in Section 4.3.1. The kernel trick for transforming non-linearly separable data into linear separable data is discussed in Section 4.3.2. Handling completely non-separable data is discussed in Section 4.3.3.

4.3.1 Linearly separable data

The foundation of support vector machine theory is built upon the concept of a *hyperplane*. In an n-dimensional inner product space, G, a hyperplane is the set, S, of points $\mathbf{x}$ satisfying

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \tag{4.3.1.1}$$

for some pair $(\mathbf{w}, b) \in G \times \mathbb{R}$. Here n may be finite or infinite and each pair $(\mathbf{w}, b)$ specifies a unique hyperplane in G. The SVM constructs a hyperplane which separates the data in such a way that all points in one class satisfy $\mathbf{w} \cdot \mathbf{x} + b > 0$ while all points in the other class satisfy $\mathbf{w} \cdot \mathbf{x} + b < 0$. Thus the hyperplane specified by $(\mathbf{w}, b)$ can be viewed as a surface that separates two classes and is called a *separating hyperplane*. An example of an optimal separating hyperplane can be seen in Figure 4.8.

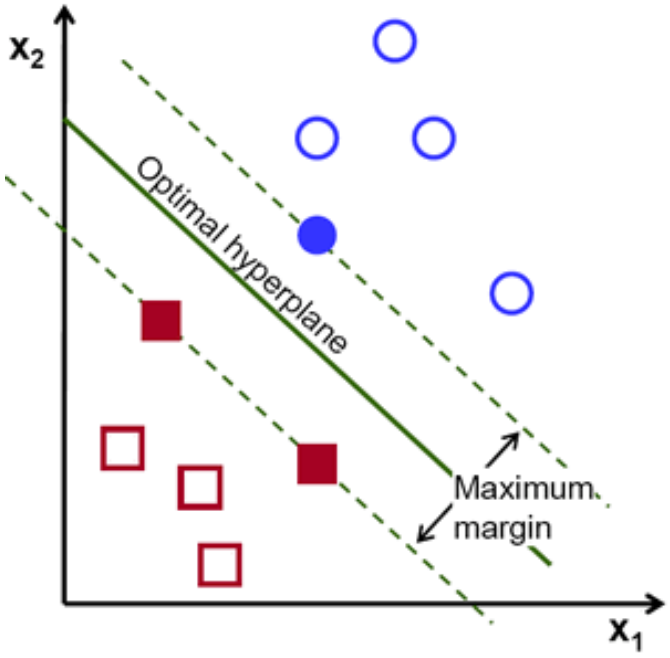


Figure 4.8: An optimal separating hyperplane between a linearly separable two dimensional data points.
© OpenCV Dev Team (2014). *Optimal Hyperplane*. Retrieved 21 October 2014, from OpenCV, Introduction to Support Vector Machines, docs.opencv.org.

More formally, consider the training set $\{\mathbf{x}_i, y_i\}, i = 1, \ldots, l$ where $y_i \in \{-1, 1\}$ is the class label for $\mathbf{x}_i \in \mathbb{R}^N$. Assume this training set is linearly separable i.e. that a separating hyperplane exists that splits the positive from the negative examples. Thus, a $\mathbf{w} \in \mathbb{R}^N$ and $b \in \mathbb{R}$ exist such that

$$\mathbf{x}_i \cdot \mathbf{w} + b \geq +1, y_i = +1 \tag{4.3.1.2}$$

$$\mathbf{x}_i \cdot \mathbf{w} + b \leq -1, y_i = -1 \tag{4.3.1.3}$$

which can be combined into one inequality:

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1 \tag{4.3.1.4}$$

with equality occurring in the case of at least one $\mathbf{x}_i$. The equality follows since $(\mathbf{w}, b)$ can be scaled without changing the hyperplane. The set of hyperplanes $\mathbf{w} \cdot \mathbf{x} + b = 0$ that satisfy Equation 4.3.1.4 is the same as the set of all separating hyperplanes, but multiple representations of the same hyperplanes are normalised to a single representation of the hyperplane in canonical form.

The next step is to select the canonical separating hyperplane for which the closest data point $\mathbf{x}_i$ is furthest from the hyperplane. The distance of a point from the plane $\mathbf{w} \cdot \mathbf{x} + b = 0$ is $\frac{|\mathbf{w} \cdot \mathbf{x} + b|}{\|\mathbf{w}\|}$. Thus from Equation 4.3.1.4 for all the canonical separating hyperplanes, the minimum of the numerator is 1. The problem reduces to maximising $\frac{1}{\|\mathbf{w}\|}$, otherwise known as the margin. $\|\mathbf{w}\|$ is also a measure of smoothness [Girosi *et al.* 1995] or stability of the function $\mathbf{w} \cdot \mathbf{x} + b$ *i.e.* the smaller $\|\mathbf{w}\|$ is, the less $\mathbf{w} \cdot \mathbf{x} + b$ changes with a fixed change in $\mathbf{x}$. Thus, this problem can be viewed as finding the canonical separating hyperplane exhibiting the least variance.

Maximising the margin $\frac{1}{\|\mathbf{w}\|}$ is equivalent to minimising $\|\mathbf{w}\|$ or $\frac{1}{2}\|\mathbf{w}\|^2 = \frac{1}{2}\mathbf{w} \cdot \mathbf{w}$ which corresponds to finding the smoothest canonical separating hyperplane. This is a constrained optimisation problem:

$$\begin{aligned} \text{minimise} \quad & \frac{1}{2}\mathbf{w} \cdot \mathbf{w} \\ \text{subject to} \quad & y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \\ & i \in \{1, \ldots, l\} \end{aligned} \tag{4.3.1.5a}$$

Since the training data only appears in the form of dot products between vectors, the problem is converted to a Lagrangian formulation. Thus, the constraints in 4.3.1.4 are replaced by positive Lagrange multipliers, $\alpha_i, i = 1, \ldots, l$, one for each of the inequality constraints. This gives the *primal* Lagrangian:

$$L_P = \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}) - \sum_{i=1}^{l} \alpha_i \{y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1\} \tag{4.3.1.6}$$

According to optimisation theory, the solution to this problem lies at the saddle point of the Lagrangian function L_P which has to be minimised with respect to $\mathbf{w}$ and b, and

simultaneously maximised with respect to α where $\alpha \geq 0$. Taking partial derivatives with respect to $\mathbf{w}$ and b and setting them to zero gives

$$\frac{\partial}{\partial \mathbf{w}} L_P = \mathbf{w} - \sum_{i=1}^{l} \alpha_i y_i \mathbf{x}_i = 0 \tag{4.3.1.7a}$$

$$\frac{\partial}{\partial b} L_P = -\sum_{i=1}^{l} \alpha_i y_i = 0 \tag{4.3.1.7b}$$

Equation 4.3.1.7a shows that $\mathbf{w}$ is a linear combination of the $\mathbf{x}_i$'s while 4.3.1.7b indicates that the sums of coefficients of data points from each class are equal. Furthermore, the Karush-Kuhn-Tucker (KKT) complementarity conditions between the α_i and their corresponding constraints can be exploited to find out more about the linear combination. The relevant KKT complementarity conditions in this case which hold at the solution are

$$\alpha_i \{y_i(\mathbf{w} \cdot \mathbf{x} + b) - 1\} = 0, i \in \{1, \ldots, l\} \tag{4.3.1.8}$$

Therefore $\alpha_i = 0$ or $y_i(\mathbf{w} \cdot \mathbf{x} + b) - 1 = 0$. This implies that only the data points x_i closest to the hyperplane at a distance of $\frac{1}{\|\mathbf{w}\|}$ can have non-zero Lagrange multipliers α_i. These points, called *support vectors*, usually form a small proportion of the data [Boser *et al.* 1992], and so SVMs are said to learn a decision function with a sparse representation [Brown *et al.* 1999; Campbell 2002] and are used to calculate $\mathbf{w}$ from Equation 4.3.1.7a.

Substituting Equations 4.3.1.7a and 4.3.1.7b into 4.3.1.6, the *Wolfe Dual* problem of maximising

$$L_D = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \tag{4.3.1.9}$$

with respect to α subject to $\alpha \geq 0$ and Equation 4.3.1.7b is obtained [Fletcher 1987]. This problem is equivalent to the original primal optimisation problem in 4.3.1.6 as the α obtained is identical. The expression in 4.3.1.9 for L_D is a quadratic form, and the constraints are all linear in the α_i's. Thus, this is a standard quadratics programming problem and can be optimised by various computer packages.

Suppose an α^0 which maximises L_D subject to the constraints have been found. Let $\mathbf{w}^0$ and b^0 be the parameters of the corresponding optimal hyperplane. From Equation 4.3.1.7a, $w^0 = \sum_{i=1}^{l} \alpha^0 y_i x_i$. To calculate b^0 the KKT complementarity conditions (4.3.1.8) must be used. For any non-zero α_i^0, it follows that

$$y_i(\mathbf{w}^0 \cdot \mathbf{x}_i + b^0) = 1 \tag{4.3.1.10}$$

so that

$$b^0 = y_i - \mathbf{w}^0 \cdot \mathbf{x}_i \tag{4.3.1.11}$$

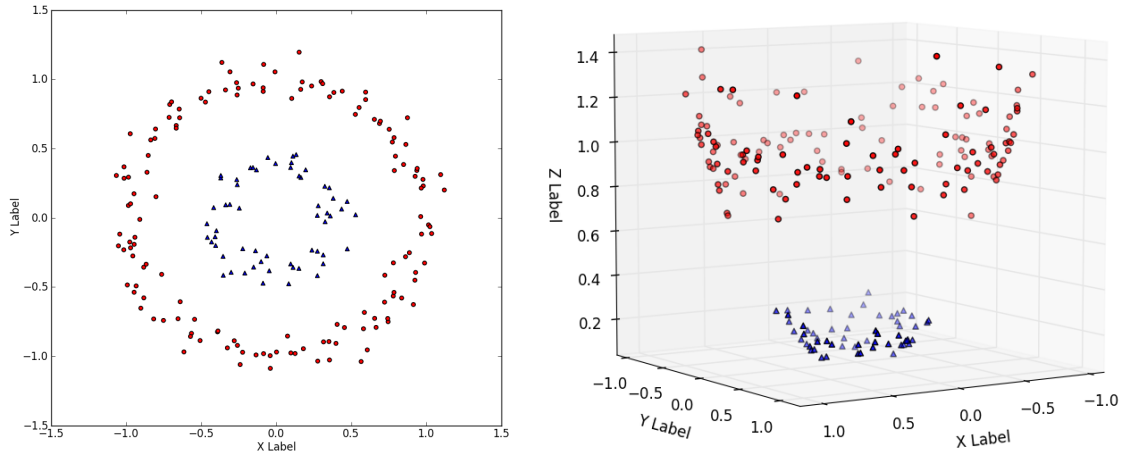
since $y_i = \pm 1$.

Once the separating hyperplane has been determined, it can be used to classify new data points by considering which side of the hyperplane they lie on. Thus the decision or classification function for a new point $\mathbf{x} \in G$ is

$$sgn(\mathbf{w}^0 \cdot \mathbf{x} + b^0) = sgn(\sum_{i=1}^{l} \alpha_i^0 y_i \mathbf{x}_i \cdot \mathbf{x} + b^0)$$

4.3.2 Kernel trick

The motivation for the kernel trick is the case where the data is not linearly separable in the input space G. Cover's Theorem gives us that this data can usually be transformed into a higher dimensional space in which it is linearly separable [Haykin 1998]. An example of non-linearly separable data and its corresponding transformation can be seen in Figure 4.9.



(a) (b)

Figure 4.9: A data set in $\mathbb{R}^2$ (4.9a) and its corresponding transformation by the function $f(x, y) = x^2 + y^2$ (4.9b) in which it is linearly separable.
Images courtesy of Eric Kim (2013). Retrieved 21 October 2014, from Everything You Wanted to Know about the Kernel Trick, www.eric-kim.net.

The kernel trick derives from Mercer's theorem which states that for any bounded, non-negative function $L(x, y)$ which is symmetric in it's arguments $x, y \in G$, there exists a Hilbert space H_K, called the *feature space*, and a mapping $\phi : G \to H_K$ such that $\phi(x) \cdot \phi(y) = K(x, y)$ for all $(x, y) \in G^2$ [Kroon 2003]. Such a K is known as the *kernel function*. If ϕ maps the input space into a higher-dimensional space, the transformed data points, $\phi(x_1), \ldots, \phi(x_l)$), are more likely to be linearly separable than the original data points.

Thus, to generate a separating hyperplane in the higher dimensional space, the method described in Section 4.3.1 is repeated on the transformed data points, $\phi(x_1), \ldots, \phi(x_l)$). This corresponds to a non-linear separating hyperplane in the original input space. In order to generate this non-linear separating hyperplance, we must maximise

$$L_D = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i \cdot \mathbf{x}_j) \qquad (4.3.2.1)$$

subject to $\alpha \geq 0$ and $\sum_{i=1}^{l} \alpha_i y_i = 0$. This is identical to Equation 4.3.1.9 except $\mathbf{x}_i \cdot \mathbf{x}_j$ has been replaced by $K(\mathbf{x}_i \cdot \mathbf{x}_j)$. The mapping ϕ was removed by replacing every expression of the form $\phi(x) \cdot \phi(y)$ for $(x, y) \in G^2$ with $K(x, y)$ since the optimisation problem in Equation 4.3.1.9 for finding the separating hyperplane only made use of the data points in the form of inner products. Calculating the inner product of the transformed data points directly is known as the *kernel trick*.

Since the higher-dimensional space H_K associated with K is usually infinite dimensional and ϕ is unknown, the explicit calculation of $w^0 = \sum_{i=1}^{l} \alpha_i^0 y_i \alpha(\mathbf{x}_i)$ difficult. Fortunately, it is the decision function $sgn(w^0 \cdot \phi(\mathbf{x}) + b^0)$ for $\mathbf{x} \in G$ that is of interest. Substituting the expression for w^0 into this decision function yields

$$\begin{aligned} & sgn((\sum_{i=1}^{l} \alpha_i^0 y_i \phi(\mathbf{x}_i)) \cdot \phi(\mathbf{x}) + b^0) \\ =& sgn(\sum_{i=1}^{l} \alpha_i^0 y_i K(\mathbf{x}_i, \mathbf{x}) + b^0) \end{aligned}$$

where the kernel trick was used to replaced the inner product in the higher-dimensional space. The value of b^0 is the mean (over nonzero α_i^0) of the values obtained from

$$\begin{aligned} b^0 &= y_i - w^0 \cdot \phi(\mathbf{x}_i) \\ &= y_i - \sum_{i=1}^{l} \alpha_j^0 y_j K(\mathbf{x}_j, \mathbf{x}_i) \end{aligned}$$

Thus with the kernel trick, the optimisation problem can be solved in the higher dimensional space, calculate b, and implicitly define the hyperplane without ever having to directly transform the data points. An example of this hyperplane can be seen in Figure 4.10.

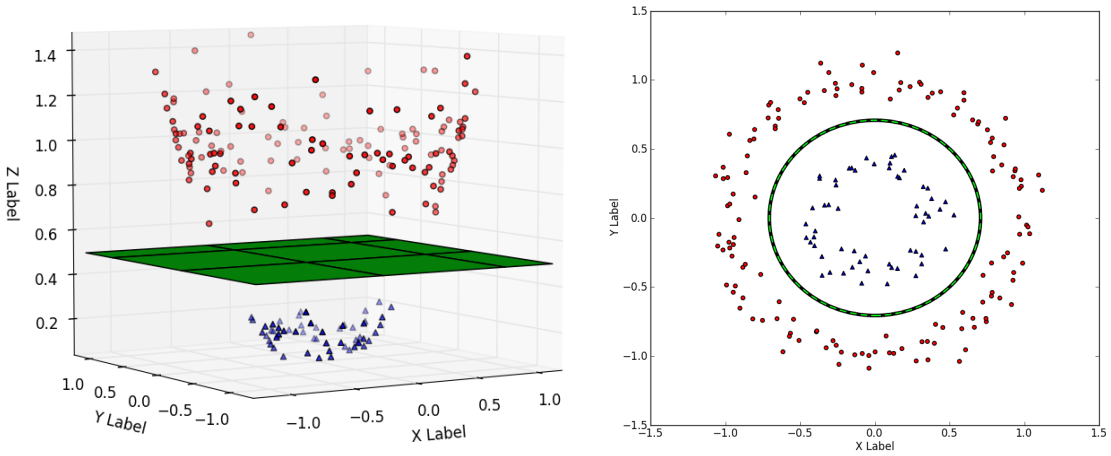
Some common kernel functions are listed below [Hsu *et al.* 2003]:

- linear: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$
- polynomial: $K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d, \gamma > 0$
- radial basis function (RBF): $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma ||\mathbf{x}_i - \mathbf{x}_j||^2), \gamma > 0$
- sigmoid: $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i, \mathbf{x}_j + r)$

Here γ, r, and d are kernel parameters.

4.3.3 Inseparable data

The SVM also handles the case where the data is inseparable, even in a higher-dimensional space, by means of a *soft margin* hyperplane [Cortes and Vapnik 1995]. The approach of the soft margin SVM is to allow the constraints in Equation 4.3.1.4 to be violated but in such a way that will increase the objective function. Thus, the margin $\frac{1}{||w||}$ is now



(a) (b)

Figure 4.10: The linear hyperplane in the higher-dimensional space (4.10a) with its corresponding non-linear hyperplane in the original input space (4.10b).
Images courtesy of Eric Kim (2013). Retrieved 21 October 2014, from Everything You Wanted to Know about the Kernel Trick, www.eric-kim.net.

interpreted as the distance which must be attained between the separating hyperplane and a point in order to avoid penalisation of the objective. This is done by introducing slack variables $\xi = (\xi_1, \ldots, \xi_l)$ which modifies the constraints to

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) \geq 1 - \xi_i, i \in 1, \ldots, l \tag{4.3.3.1}$$

for $\xi \geq 0$ where ξ_i is as small as possible. The hard-margin objective function is thus changed to

$$\frac{1}{2}\mathbf{w} \cdot \mathbf{w} + C \sum_{i=1}^{l} \xi_i \tag{4.3.3.2}$$

where $C \in \mathbb{R}$ is a parameter determining the extent to which violations of the original constraints in Equation 4.3.1.4 should be penalized.

The quadratic programming is similar to that in Section 4.3.1. Lagrange multipliers, $\alpha \geq 0$ and $\eta = (\eta_1, \ldots, \eta_l)$, are introduced to handle the constraint in Equation 4.3.3.2 and $\xi \geq 0$ respectively. The primal Lagrangian is thus

$$L_P(w, b, \eta, \alpha, \eta) = \frac{1}{2}\mathbf{w} \cdot \mathbf{w} + C \sum_{i=1}^{l} \eta_i - \sum_{i=1}^{l} \alpha_i \{y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 + \eta_i\} - \sum_{i=1}^{l} \eta_i \xi_i \tag{4.3.3.3}$$

Taking partial derivatives with respect to w, b, ξ_i, and setting them to zero we obtain the same constraints as before,

$$w = \sum_{i=1}^{l} \alpha_i y_i x_i \tag{4.3.3.4}$$

and

$$\sum_{i=1}^{l} \alpha_i y_i = 0 \tag{4.3.3.5}$$

along with a new constraint

$$C = \alpha_i + \eta_i \tag{4.3.3.6}$$

These constraints yield the Wolfe dual objective function

$$L_D(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \tag{4.3.3.7}$$

Thus the modification of the objective function has no effect on the dual objective function but there is a change in the constraints. For the separable case, $\alpha \geq 0$ and $\sum_{i=1}^{l} \alpha_i y_i = 0$. In the current problem, $\alpha_i \leq C$, which follows from Equation 4.3.3.6 and $\eta \geq 0$. The modification for linear inseparability has the effect of limiting the search space for α.

Once an optimal α^0 is found, the corresponding w^0 and b^0 can be found as in Section 4.3.1.

4.4 Conclusion

In this chapter, we discussed the algorithms that are the foundation of our research, namely the Viola Jones object detection framework and the Support Vector Machine. In the next chapter, we discuss how they were implemented and used to carry out our research.

Chapter 5

Research Methodology

5.1 Introduction

In this chapter, we discuss how the algorithms discussed in the previous chapter were applied and tested on our rock art dataset. We first outline the aims and scope of our research in Section 5.2 and 5.3 respectively. The open source implementations for our selected algorithms are listed and discussed in Section 5.4. In Section 5.5, we list our research questions and in Section 5.6, we describe the methodology employed to answer those questions.

5.2 Aims

The primary aim of this research was to investigate the southern African rock art domain and select object detection techniques to identify and label the numerous objects and figures that appear in the South African Rock Art Digital Archive. This would open up the archive by allowing text based search queries allowing retrieval of images according to object by rock art researchers.

5.3 Scope

We narrowed the scope of our research to fit time constraints and the limitations of our object detection algorithms but were careful these would not diminish the value of the results obtained.

Collecting positive samples for each type of object was a difficult and time consuming task as it required searching through the SARADA and manually cropping out each object. Collecting negative samples of non-object regions was slightly easier but still a time consuming task. Even more time was required to train each classifier, with the training times taking between two and six weeks depending on the final number of samples in the dataset. Clearly, the number of objects we could create classifiers for would be limited by the time available to carry out this research. Given this constraint, we decided that training classifiers for four objects would provide a reasonable estimation on the performance and utility of the Viola Jones object detection framework and SVM combination. The criteria used to select these four objects were that there had to be at least 400 instances of

Object	Number of images	Number of objects
Humans	1358	4931
Humans elong.	523	1912
Eland	1136	1817
Horses	365	1113
Cattle	324	1101
Rhebuck	365	626
Elephants	331	443
Felines	323	365
Baboons	150	267
Snakes	246	258
Birds	77	175
Rhinos	131	145
Rain animals	136	143
Hippos	41	53
Springbuck	21	47
Bee swarms	32	42
Crabs	5	6
Jackals	2	2
Total	5566	13446

Table 5.1: Number of images and objects in each object class.

the object (but preferably more) to provide a reasonable sample spread. We also wanted to maximise the inter-class variation between the four objects to see how our algorithms would handle different looking objects. As can be seen in Table 5.1, eland and human figures were by far the most common and so they were selected. Even though the horse and cattle objects were the next most commonly depicted in the rock art, they were too similar to each other and to eland and so were not selected. Rhebuck and elephants were the next most common figure and sufficiently different from the eland and human figures and so were also selected.

The rotation of the objects in the test images were limited to a maximum of $\pm 15^\circ$ to the horizontal as this is the maximal rotational invariance of the Viola Jones algorithm [Jones and Viola 2003]. As discussed in Section 2.7, the objects can appear at arbitrary rotations but as Table 5.2 shows, most of our selected objects fall within the $\pm 15^\circ$ range.

Object	Total	Flat	%
Eland	1690	1313	77.69
Elephants	443	418	94.36
Humans	1080	957	88.61
Rhebuck	626	499	79.12

Table 5.2: Number and percentage of *flat* objects of the total *i.e.* objects that are rotated to within $\pm 15^\circ$ to the horizontal in each of our training datasets.

The posture of the objects was also restricted. This is because each posture changes the fundamental shape of the object and would require a different classifier to be trained. Therefore, we selected the most common postures for each object which are the standing/walking side profile views.

5.4 Programs and tools

The following programs and tools were utilised to aid this research and for their implementations of the algorithms described in Chapter 4.

5.4.1 ImageClipper

ImageClipper [Seo 2008] is a free and open source tool developed for computer vision applications. It can manually crop, rotate, and scale images and video sequences frame by frame to create training sets. Information such as the location and rotation is also automatically stored allowing it to be used as a ground truth for testing experiments.

5.4.2 OpenCV

Open Source Computer Vision, or OpenCV for short [Intel *et al.* 2014], is a free and widely used computer vision and image processing application. Among other libraries and routines, it has a built-in implementation of the Viola Jones object detection framework.

5.4.3 LIBSVM

LIBSVM [Chang and Lin 2011] is an implementation for support vector machines with automatic parameter selection and supports each of the kernels described in Section 4.3.2.

5.5 Research questions

The scope of this research can be formulated by listing the following research questions:

1. Can the Viola Jones object detection framework trained with Haar-like features detect and recognise multiple rock art objects in full raw photographs?

2. Can a secondary classifier, such as an SVM, which is passed the sub-windows detected by the Viola Jones classifier, improve the classification performance?

5.6 Methodology

5.6.1 Data collection and processing

Data collection

As discussed in Section 2.2.2, a small subset of images in the SARADA have been manually labelled. For this research, we were interested in the animal and human subclasses and so these images were downloaded and organised into the subclasses listed in Table 5.1.

Training set creation

The positive training sets for each of our selected objects were manually cropped out of the source images. Since the general shape of an object is different depending on its rotation and the direction its facing, it would confuse the training process. Therefore, the sets were normalised by rotating each object until they were upright and flipped horizontally, if necessary, so that the head faced to the right. Examples of the result of this step for the eland classifier can be seen in Figure 5.1.



(a) RARI-LEE-RSA-CYH1-3



(b) RARI-RSA-GAM1-56



(c) RSA-BAE2-6



(d) RSA-POE1-1

Figure 5.1: Cropped image samples from the positive training set for eland with 5.1a and 5.1c also flipped horizontally.
Images reproduced with permission from the Rock Art Research Institute, University of the Witwatersrand, South Africa. Retrieved 26 October 2014, from SARADA, www.sarada.co.za.

Since the Viola Jones classifier search window is at a fixed height to width ratio, the average width, height, and height to width ratio of all the positive images in each class were calculated and can be seen in Table 5.3.

Object	Height(px)	Width(px)	Height:Width
Eland	45	80	1:1.78
Elephants	50	70	1:1.39
Humans Upright	84	40	1:0.47
Rhebuck	50	76	1:1.52

Table 5.3: Average heights and widths for each object along with the height:width ratio.

The Viola Jones algorithm also requires the negative training images to be at the same height to width ratio of the classifier window. Therefore, the height to width ratios calculated above were used to randomly crop out sub-windows from the object classes. Negative images used for the rhebuck classifier can be seen in Figure 5.2.



(a) NMSA-PJV-01-95HC



(b) RARI-HPC-01-51HC



(c) RARI-RSA-BOW1-36



(d) RARI-RSA-LOW3-32

Figure 5.2: Cropped image samples from the negative training set for rhebuck. Any non-rhebuck objects could be present in these images such as, but not limited to, humans (5.2a), eland (5.2b), a bare cave wall (5.2c), and an elephant head (5.2d).
Images reproduced with permission from the Natal Museum (Department of Archaeology & Anthropology), Pietermaritzburg, South Africa (5.2a) and the Rock Art Research Institute, University of the Witwatersrand, South Africa (5.2b, 5.2c, 5.2d). Retrieved 26 October 2014, from SARADA, www.sarada.co.za.

5.6.2 Parameter selection

The Viola Jones object detection framework, more specifically the OpenCV implementation, consists of a number of parameters that can be tuned. We experimented on a rhebuck dataset to optimise the following parameters: feature type, number of classifiers, and boosting type. Other parameters such as the positive and negative training set size were out of our control as we were already using all of the limited images available to us. The results of these experiments can be seen below.

Feature type

OpenCV supports the Haar-like features of Viola and Jones [2001], known as *basic* in OpenCV and can be seen in Figure 4.1, and Lienhart *et al.* [2003], known as *all* in OpenCV and can be seen in Figure 4.4. The basic features had a superior true positive rate of 80% compared to 70% for all features. All features were still selected since the false positives per image was 1.14 which is less than half that of the basic features.

Feature type	True positive rate (%)	False positives per image
Basic	80	2.34
All	70	1.14

Table 5.4: Performance comparison of the Haar-like feature types.

Number of classifiers

The number of classifiers in the Viola Jones cascade that we tested had 5, 10, 15, and 20 classifiers. Table 5.5 shows that the true positive rate performance increases from 20% with 5 classifiers to 93% with 15 classifiers then decreases to 70% with 20 classifiers. The false positives per image decreases from 11.39 with 5 classifiers to 1.14 with 20 classifiers. Even though the cascade with 15 classifiers had a superior true positive rate of 93% compared to the 70% of the cascade with 20 classifiers, we still selected the cascade with 20 classifiers since it had a lower amount of false positives per image of 1.14 compared to 2.06. In Figure 5.3 we calculated the precision on a hypothetical dataset of 100 images using the true positive rates and false positives per image from Table 5.5. We can see that no matter how many objects in a dataset, the precision would still be higher with the 20 classifier cascade. Precision was important considering that only a small subset of the SARADA contains each type of object.

Classifiers	True positive rate (%)	False positives per image
5	20	11.39
10	59	9.68
15	93	2.06
20	70	1.14

Table 5.5: Performance comparison of different numbers of classifiers in the cascade.

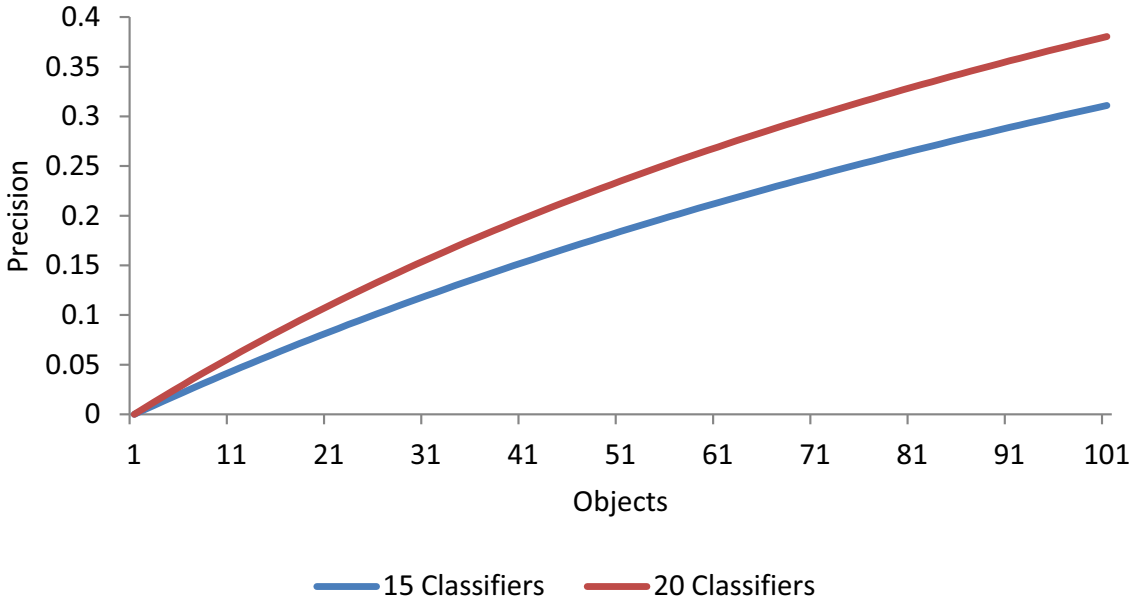


Figure 5.3: Comparison of precision of the Viola Jones cascades with 15 and 20 classifiers respectively.

Boosting type

OpenCV supports four variants of the AdaBoost algorithm, Gentle, Logit, Real, and Discrete. As can be seen in Table 5.6, the classifier trained with Real AdaBoost outperformed the others with a true positive rate of 82% however there were 2.47 false detections per image. Gentle and Discrete AdaBoost performed the next best with a true positive rate of 70% although Gentle Adaboost only had 1.14 false positives per image compared to the 1.82 or Discrete AdaBoost. LogitBoost performed the worst with only a 14% true positive rate and 1.11 false positives per image. Therefore, we selected Gentle AdaBoost since it had the best balance between the positive rate and false positives per image. The reason for this slightly better performance was because each classifier was trained using weighted least-squares regression instead of performing exact optimisation as with Real AdaBoost.

Boosting type	True positive rate (%)	False positives per image
Real AdaBoost	82	2.47
Gentle AdaBoost	70	1.14
Discrete AdaBoost	70	1.82
LogitBoost	14	1.11

Table 5.6: Performance comparison of the different boosting types.

SVM feature type

The feature vector selected for the SVM was a one dimensional vector consisting of all the pixel intensity values of each image. The images were normalised by resizing each image to a fixed size and only consisted of a single object.

5.6.3 Experimental setup

Five-fold cross validation was selected as the experimental procedure for the Viola Jones classifier. The original source images were used in the positive and negative testing sets (instead of the cropped images created for training as described in Section 5.6.1). As discussed in the Scope (5.3), the maximal rotational tolerance of the Viola Jones algorithm was found to be $\pm 15°$ and so the test images with any object beyond this threshold were discarded. Otherwise, the test images fairly represented the characteristics of southern African rock art as discussed in Chapter 2. Objects from all the different sub-classes in Table 5.1 were present in the positive and negative testing sets. The objects could face in either direction, overlap, and appear deteriorated. The average sizes of the datasets can be seen in Table 5.7.

Object	# + Train	# + Test	# - Train	# - Test
Eland	1452.8	210.8	1751	114
Elephant	323.2	44.8	2663	80
Humans Upright	753	148	2712	100
Rhebuck	466.6	45.6	2604	80

Table 5.7: Average sizes of the positive/negative training/testing datasets used in the five-fold cross validation.

The SVM was also trained with five-fold cross validation. The same training and testing sets were used except the testing sets here were not the raw original source images but contained exactly exactly one of the desired object or another object entirely. This was acceptable since it was simulating the sub-windows that the Viola Jones classifier would pass to it.

5.7 Conclusion

At the beginning of this chapter we discussed the aims and scope of this research and learnt how it was narrowed to creating classifiers for just four objects with a limited rotation. The motivations for selecting the algorithms for this research were described and we saw that standard segmentation and smoothing algorithms fail or cannot be generalised when applied to rock art. We also listed programs and tools that implement our selected algorithms. The research questions were formulated and the methodology for answering them were detailed. This included creating training sets, selecting parameters and features for the Viola Jones object detection framework and the SVM, and conducting the training and testing using five-fold cross validation. The results of this process is given in the next chapter.

Chapter 6

Results

6.1 Introduction

The performances of our classifiers are presented in this chapter along with its analysis and discussion. The performance of the Viola Jones classifier is presented in Section 6.2. while the performance of the SVM is discussed in Section 6.3. The performance of the cascade when applied to an unlabelled portion of the SARADA is then presented in Section 6.4.

6.2 Analysis of Viola Jones classifier

Figures 6.1 and 6.3, show that the four Viola Jones classifiers correctly located and classified on average 59.44% of their objects at a false positive rate of 0.96 per image. To fully understand these figures and put them into context, we individually analyse the true positive, false negative, false positive, and true negative rates of each of the four classifiers and from there gain a more complete view of the performances.

6.2.1 True positive analysis

The true positive rate is the percentage of objects located and correctly classified by their respective classifiers.

Figure 6.1 shows that the average true positive rate of the four classifiers for eland, elephants, humans, and rhebuck was 59.44%. Usually these detections were of paintings that were still largely intact that did not display many of the characteristics of deterioration discussed in 2.8. On closer inspection, we can see that the eland, human, and rhebuck classifiers had similar true positive rates of 60.92%, 63.45%, and 66.7% respectively while the elephant classifier performed relatively poorly with a true positive rate of 46.68%.

The performance of the elephant classifier can be attributed to the lower number of positive training samples which, from Table 5.7, stood at 323.2 samples on average from the five cross-validation sets compared to 1452.8, 753, and 466.6 for eland, humans, and rhebuck respectively. Plotting the positive training sample set size against against the true positive rate in Figure 6.2, we notice a sharp drop in performance between 466.6

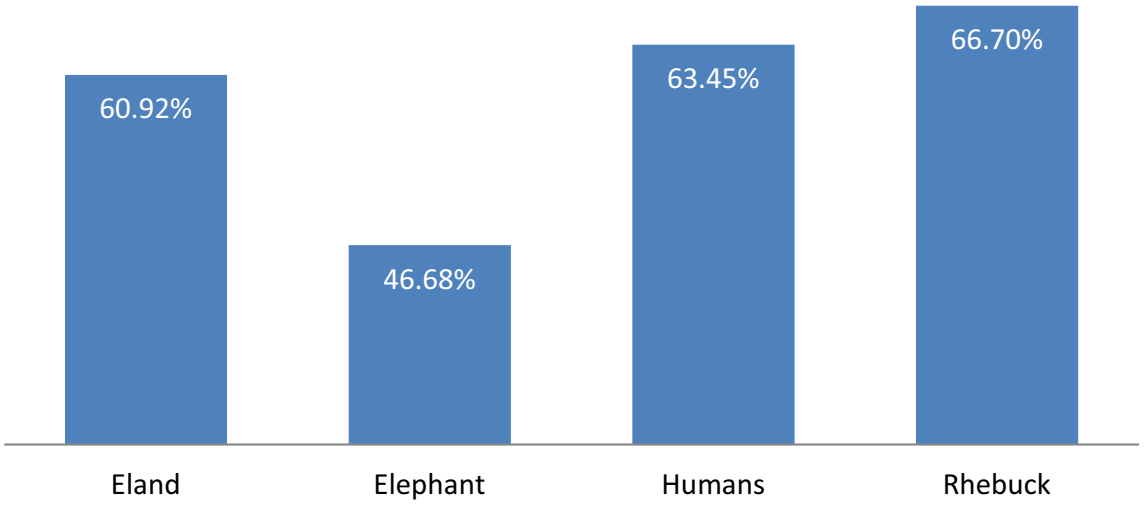


Figure 6.1: Viola Jones classifiers true positive rate summary. The average rate between the four classifiers is 59.44%.

and 323.2 samples. This seems to suggest that that there is a threshold between these two values where the performance of the Viola Jones object detection framework breaks down on rock art images. This could be because the complexity of the samples increases with the size.

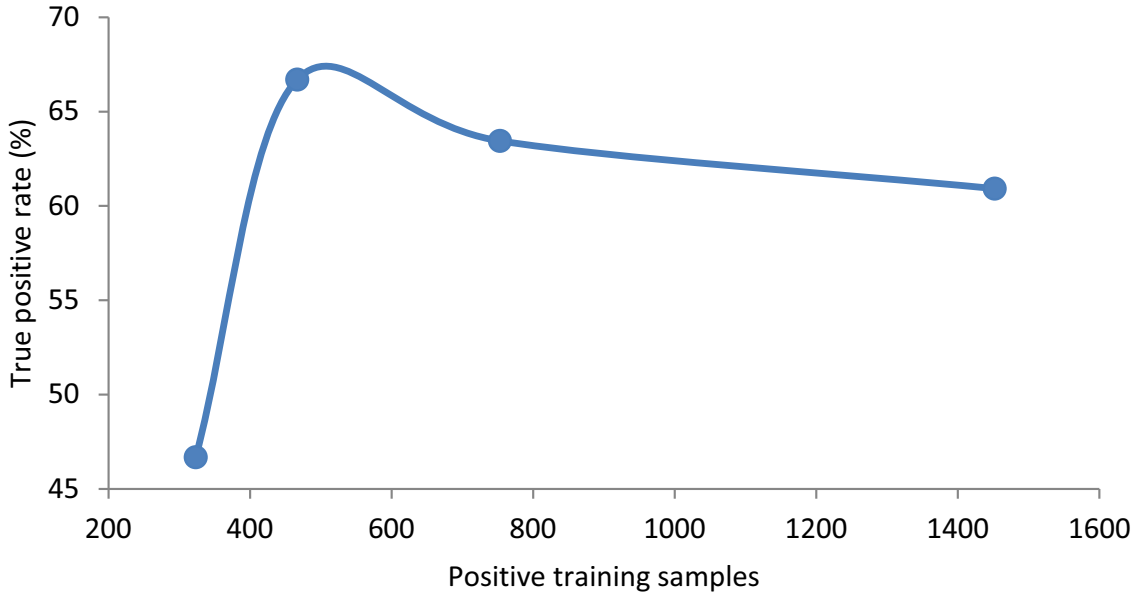


Figure 6.2: Training set size vs. true positive rate of the Viola Jones classifiers.

Conversely, increasing the positive training sample set size indefinitely does not seem to correlate with an increase in true positive rate. Instead, as Figure 6.2 shows, the true positive rate seems to peak at 67% with approximately 500 positive training samples and then level off (albeit at a slight downward slope) at about 61%. Therefore, the optimal positive training sample set size seems to fall between 500 and 700.

The correlation between the positive training sample set size and the true positive rate must not be looked at too closely as there are other factors at play that are difficult to measure. By inspection of the datasets, we noticed that was a greater variation in the eland and human datasets compared to the elephant and rhebuck datasets. Eland and humans are the most commonly depicted objects in the rock art and so naturally there was a greater variation in the proportions of their features. We can infer from this that the classifier may have begun to over generalize which would explain the slight decrease in performance when compared to rhebuck.

The relatively poor performance of the elephant classifier could also be explained by the proportionately higher number of engravings and sketches in its dataset while the other three contained more paintings. It is likely that the Viola Jones algorithm was mistakenly selecting background features inside and around the outlines of the objects resulting in a less effective classifier.

Examples of true positive detections for each of the eland, elephants, humans, and rhebuck can be seen in Figures C.1 to C.4.

6.2.2 False negative analysis

The false negative rate is the percentage of objects missed by their respectively classifiers and incorrectly classified as a background object.

An analysis of these figures that were not detected at all is also revealing. Since the true positive rate on average is 59.44% , the false negative rate is clearly 40.56%. These images usually consisted of objects that were highly deteriorated, painted on rough rock surfaces, or were engravings. The reasons these images were rejected lie in the calculation and selection of the Haar-like features.

Highly deteriorated images where the portions of the art are missing will most likely have the rock surface at those regions. Clearly, the Haar-like features calculated there will not correspond to those of a painting leading to the rejection of the entire sub-window containing the object. Faded paintings, on the other hand, do not offer up a high enough contrast between itself and the background rock surface. The Haar-like features were simply not sensitive enough and classified the sub-window as the background.

Rough rock surfaces slightly alter the shading of the painting or cast small shadows across it. When the Haar-like features were calculated at these unexpected regions, they rejected the current sub-window. Similarly, variably coloured rock surfaces posed the same problem.

Engravings are usually outlines of an object. Therefore, the only important features are the lines themselves with everything around them the background. This makes it difficult for the Viola Jones algorithm to select the features that encode the lines but ignore the background. From the poor performance on the engravings, we can infer that the features were mistakenly encoding the background as part of the object and so when it failed to find them in different images, the sub-windows containing the object were

rejected.

Examples of the false negative detections for eland, elephants, humans, and rhebuck can be seen in Figures D.1 to D.4.

6.2.3 False positive analysis

The false positives per image metric is the number of sub-windows in the image incorrectly classified as the object its classifier was searching for. The upper bound of this value is the number of sub-windows the classifier examines which, depending on the size of the image and classifier search sub-window, can vary from several thousand to well over one hundred thousand.

Figure 6.3 shows that the four Viola Jones classifiers had, on average, 0.96 false positives per image. These false positives were usually made up of similar looking objects. For example, Figure E.1 shows, the eland classifier confused other four legged animals, such as cattle and rhebuck, for eland.

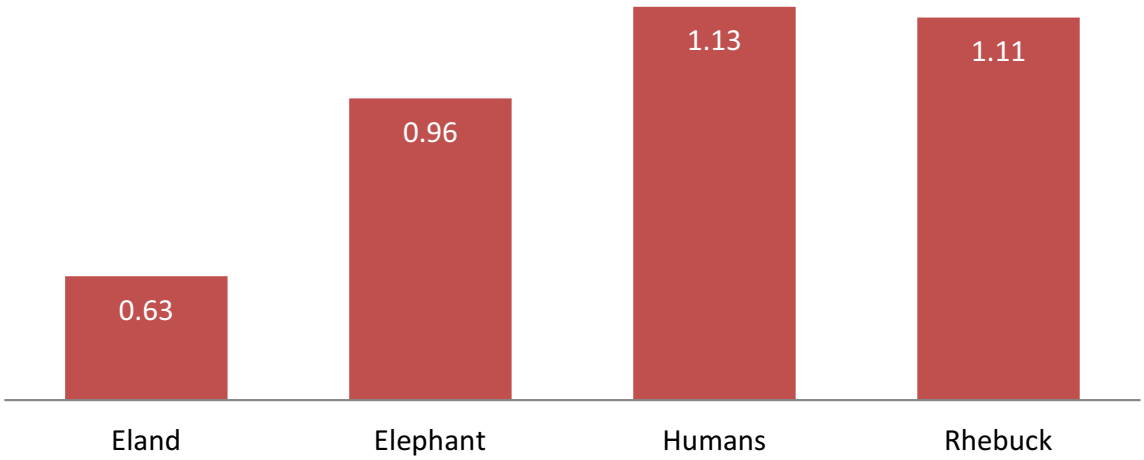


Figure 6.3: False positives per image for each of the four Viola Jones classifiers. The average amount was 0.96 false positives per image.

Looking at Figure 6.3, the eland classifier had 0.63 false positives per image which is markedly lower when compared to the other three classifiers which had just above or below 1 false positive per image. Plotting the positive training sample set size against the false positives per image, in Figure 6.4, provides insight as to why this was the case. The trend indicates that between the 323.2 positive training samples of elephant and 753 positive training samples of humans, the false positives per image seems to increase. Given that the elephant classifier only detected 46.68% of the elephants compared to the 63.45% and 66.70% of the human and rhebuck classifier respectively, it stands to reason that the elephant classifier was under trained and was less likely to detect objects, even if it was by mistake. Conversely, the human and rhebuck classifier were correctly classifying

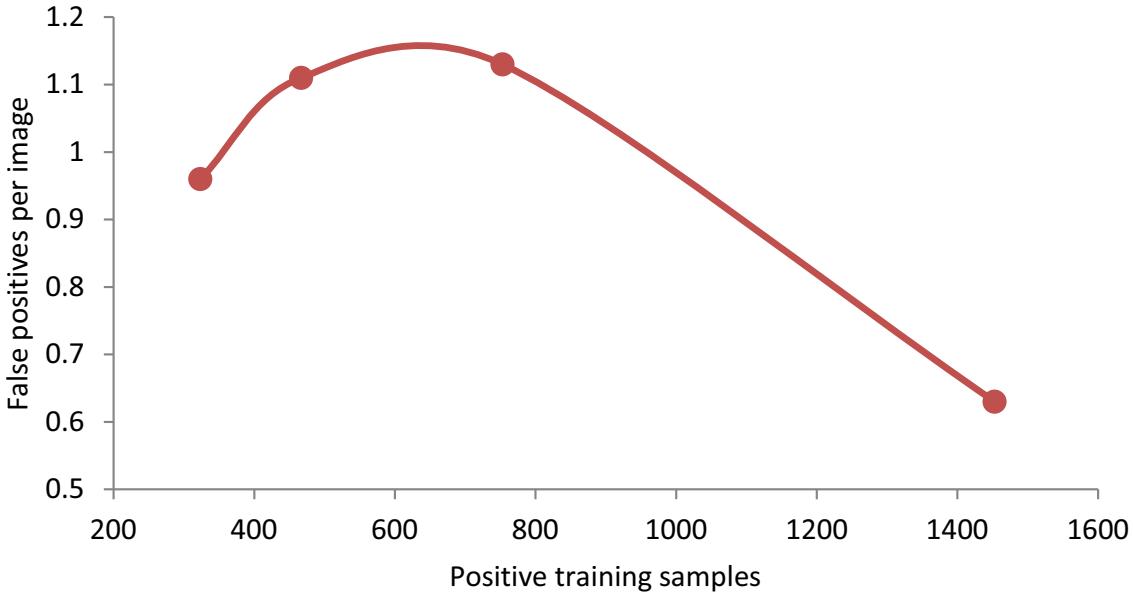
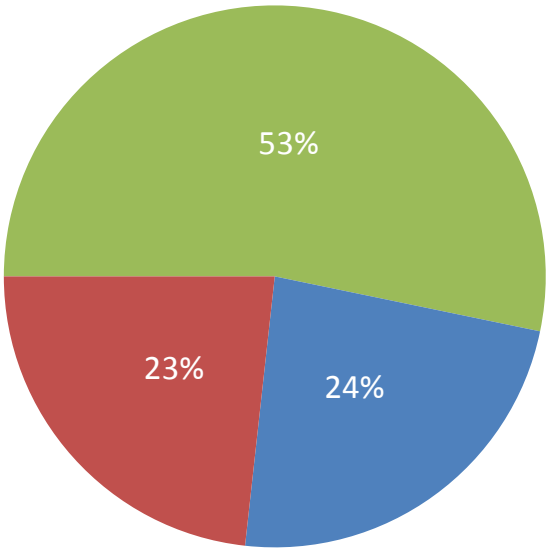


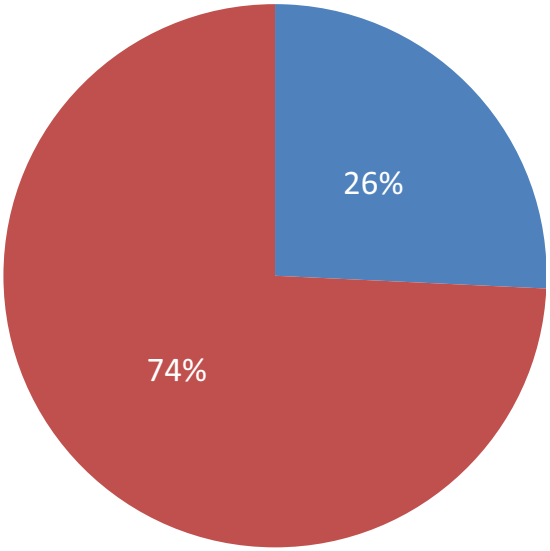
Figure 6.4: Positive training sample set size against the number of false positives per image of the four Viola Jones classifiers.

far more objects correctly but were also making slightly more mistakes. The trend then indicates that as the positive training sample set size increased to 1452.8 samples for the eland classifier, the false positive rate drops to 0.63 per image. Since the classifier had far more training samples, it was more capable of disregarding the non-eland objects but it came at the expense of a slightly lower true positive rate, compared to the human and rhebuck classifiers, of 60.92%. Given that the number of false positives per image was almost halved, while the true positive rate dropped by 3-6%, we saw this as a positive trade-off.

The average number of false positives per image of 0.96 seems to guarantee that for every image, there will be about one false positive. However, a closer inspection of the false positives indicate that this figure is actually skewed and does not give a fair reflection of the performance of the classifier. Figure 6.5a shows that 53% of images did not have any false positives while another 24% had 1 false positive per image. The other 23% of the images had at least 2 false positives per image and going up to a high of 14. Figure 6.5b shows that these 23% of images contributed 74% of the total false positives while 77% $(53\% + 24\%)$ of the images contributed to only 26% of the total false positives. It is unclear exactly why the classifiers seem to perform especially poorly on certain images while perfectly rejecting non-object regions in others but it seems it was usually when there were several similar objects in an image or even just blank walls that the classifiers seem to get completely wrong. For example, from Figure E.4c, we can see that the rhebuck classifier was detecting several deteriorated images of eland while in Figures E.4b and E.4d, the rhebuck classifier detected several sub-windows of just blank cave walls. As puzzling as this occurrence was, it is clear that there was only a proportionately small subset of images contributing to the bulk of the false positives and thus inflating the average.



(a) Images with 0 FP's. Images with 1 FP. Images with more than 1 FP.



(b) FP's from images with 1 FP. FP's from images with more than 1 FP.

Figure 6.5: False positive breakdown of the Viola Jones classifiers.

A breakdown of the false positives for each of the four classifiers can be seen in Figures B.1 to B.4.

6.2.4 True negative analysis

A true negative is a region of an image that was correctly rejected by the specific object classifier.

Since we have already discussed why objects were and were not detected in Section 6.2.1 and that it is impractical to count the number of sub-windows rejected by each classifier, we present selected images that were correctly rejected by their classifier in Figures D.1 to D.4. We examine these images for clues as to why they were rejected by looking at their features and characteristics. They show that the Viola Jones algorithm is able to sometimes differentiate between similar looking objects. Our reviewers from conferences were concerned that the Viola Jones algorithm, in an attempt to generalise to its training set, would smooth the unique features of, otherwise, similar looking objects such as quadrupeds. As discussed in the preceding section and shown in our figures, this was only true for especially difficult images.

6.3 Analysis of support vector machine classifier

The support vector machines were each tested on a positive and negative set. We analyse the true positive and false negative rates on the positive test set in Section 6.3.1 and the true negative and false positive rates on the negative test set in Section 6.3.2.

6.3.1 Performance on positive image set

Figure 6.6 shows that the support vector machine had an average true positive rate of 72.24%. Each of the individual classifiers for eland, elephants, humans, and rhebuck all had a consistent true positive rate of between 70% and 74% suggesting that the support vector machine generalises well to different objects in the SARADA.

Interestingly, the true positive rate of the elephant classifier of 71.22% was much higher than its Viola Jones counterpart which had a true positive rate of 46.68%. As mentioned in Section 6.2.1, the elephant dataset had a proportionately higher number of engravings than paintings compared to the other datasets. It would seem, then, that the support vector machine is able to separate the outlines while relegating the rock surface interior of the engraving to the background. The support vector machine also had the simpler task of performing a binary *accept* or *reject* decision on an image with exactly one elephant or none at all. However, this superior performance could adequately be utilised since the primary Viola Jones classifiers was not able to locate more potential sub-windows for the support vector machine to classify.

Also from Figure 6.6, we can see that the support vector machine had an average false negative rate of 27.76%. These images were most likely misclassified due to their deterioration. For images that were evenly faded, the pixel intensities of the image would more closely correlate to the background rock surface and so would more likely fall on the wrong side of the separating hyperplane in the higher dimensional space. Similarly, the bare rock surface would be present in deteriorated paintings where portions have flaked off and again the pixel intensities would fall on the wrong side of the separating hyperplane.

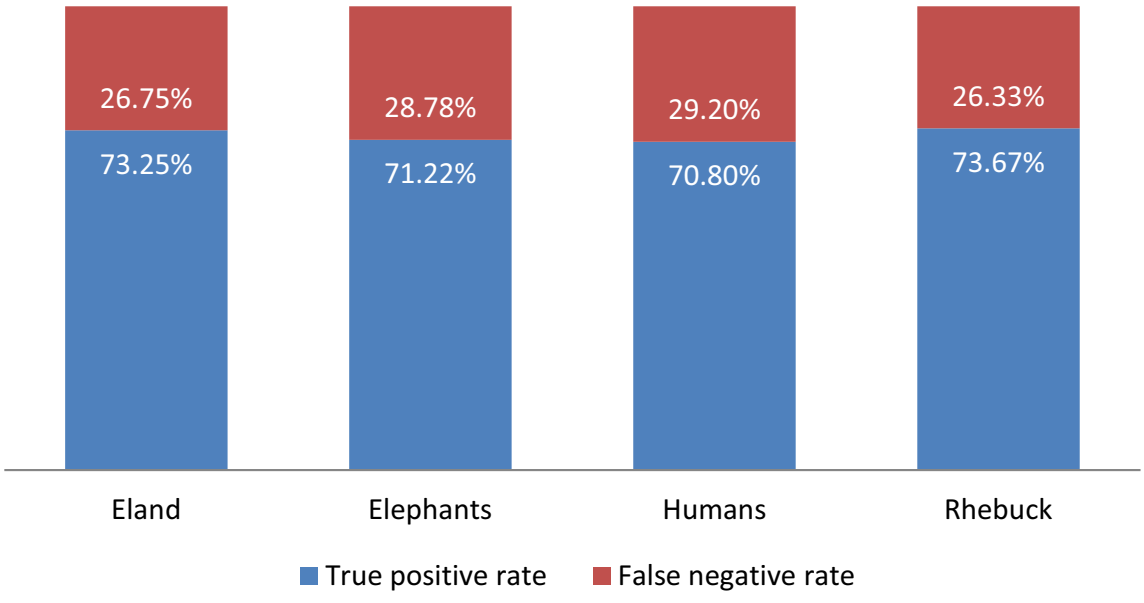


Figure 6.6: Support vector machine performance on the positive test set. The average true positive rate was 72.24% and the average false negative rate was 27.76%.

6.3.2 Performance on negative image set

Figure 6.7 shows that the support vector machine had an average true negative rate of 77.40% and an average false positive rate of 22.60%. The individual classifiers had a consistent true negative rate of between 70% and 74% with the exception of the human classifier which had a true negative rate of 94.38%. This was most likely because human figures, and thus their sub-windows, had a long height and narrow width compared to animal figures which usually had a wide width and short height. Thus, non-human figures appear in slices making their pixel intensities differ greatly to those of human figures and so would easily fall to the right side of the separating hyperplane in the higher dimensional space.

As for the other three classifiers, the false positive rate ranged from 26% to 30%. These images may have been misclassified since the other animal subclasses, especially the quadrupeds, have similar height to width ratios. When converted to grey-scale and rescaled to the dimensions of the desired object, the pixel intensities would appear similar to the desired object and so would fall to the wrong side of the separating hyperplane and be classified positively.

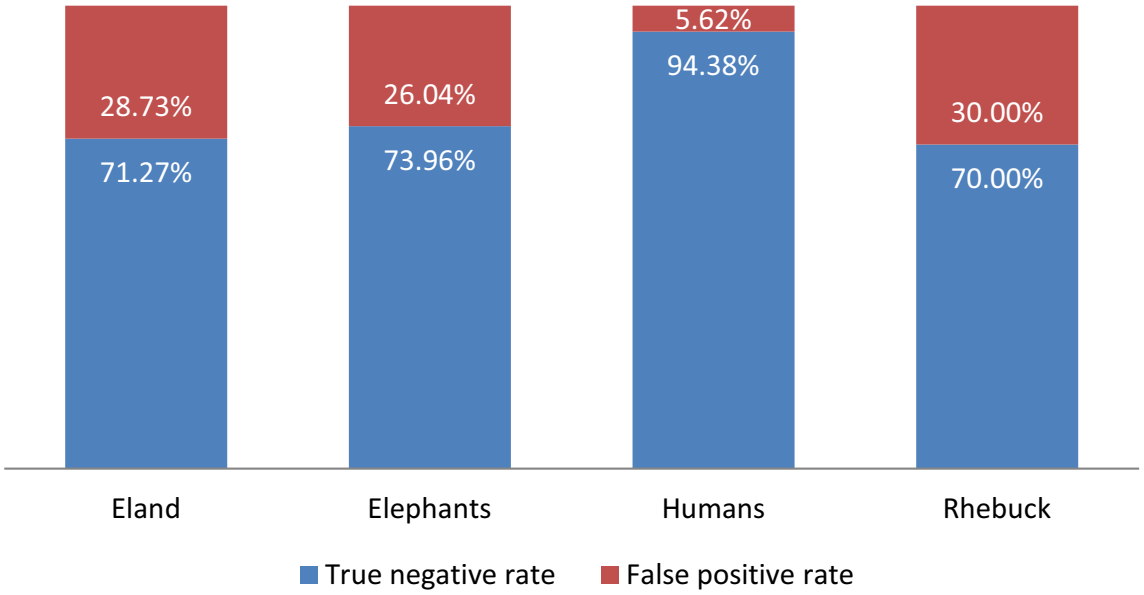


Figure 6.7: Support vector machine performance on the negative test set. The average true negative rate was 77.40% and the average false positive rate was 22.60%.

6.4 Analysis of classifiers on the SARADA

This project was commissioned with the goal of creating classifiers to detect and label the objects and figures from the entire SARADA. Therefore, in this section we gauge the performance of our classifiers on a sub-dataset of the SARADA; it is simply impractical to test all 275 000 images with no ground truths. For testing, we selected the currently

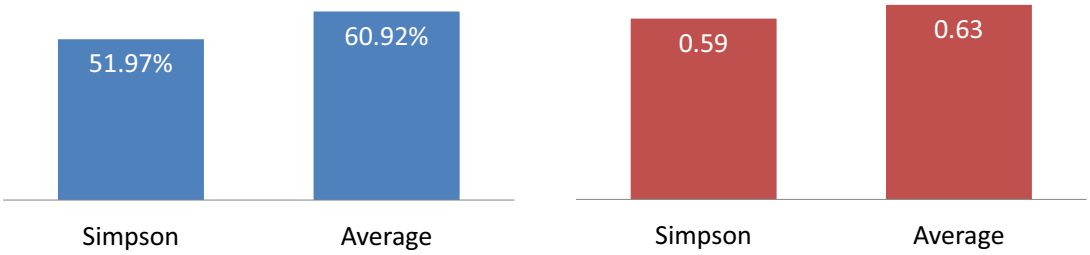
unlabelled Joan Simpson sub-dataset from the National Museum of South Africa consisting of 523 images. Since eland were the most common figure in the art, we selected the eland classifiers to run through this set.

We first ran the Viola Jones classifier through the Simpson dataset and then passed all of its detections to the support vector machine for further classification with the intention to remove as many false positives as possible. We analyse the performance of the Viola Jones classifier in Section 6.4.1, the support vector machine classifier in Section 6.4.2, and then the overall performance in Section 6.4.3.

6.4.1 Viola Jones performance

To obtain optimal results, we applied the Viola Jones classifier with the highest true positive rate from the cross-validation stage. The results obtained were consistent with the average performance of the eland classifiers during cross-validation.

From Figure 6.8a, we can see that the true positive rate obtained on the Simpson dataset was 51.97% compared to the 60.92% earlier. It must be noted that there were three near-identical outlier images that each contained nine eland, five of which were extremely faded, that were not detected. If we were to remove just two of these images, the true positive rate would increase to 58.96%, only 2% lower than the average from cross-validation.



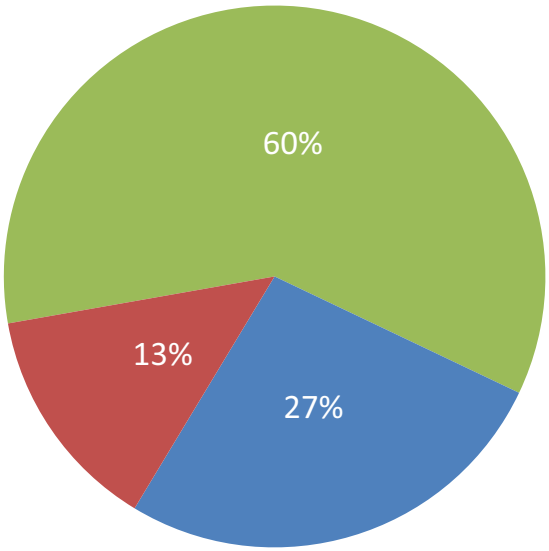
(a) True positive rate

(b) False positives per image

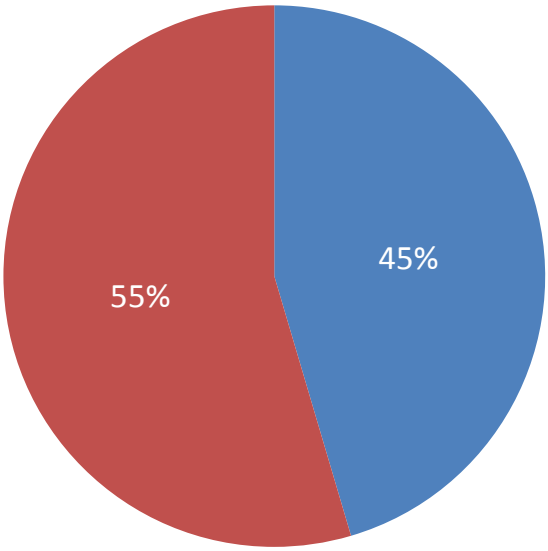
Figure 6.8: Viola Jones eland classifier true positive rate and the number of false positives per image on the Simpson dataset compared to it's average performance in cross-validation.

Figure 6.8b shows that the number of false positives per image was 0.59, performing slightly better than the average performance during cross-validation of 0.63. The false positive breakdown was also similar to the cross-validation phase. From Figures 6.9 and B.1, the percentage of images with no false positives, one false positive, and more than false positive was 60%, 27%, and 13% compared to 61%, 24%, and 15% during cross-validation. Similarly, the images with more than one false positive contributed the majority of the false detections with 55% compared to 62% during cross-validation.

There was one aspect, however, that was not consistent with the classifier performance during cross-validation. There were 83 positive images (with eland) and 440 negative images (without eland). This higher proportion of negative images resulted in the false positives dominating the total detections even though the number of false positives per image was similar to the cross-validation phase. To measure this, we introduce a new



(a) Images with 0 FP's. Images with 1 FP. Images with more than 1 FP.



(b) FP's from images with 1 FP. FP's from images with more than 1 FP.

Figure 6.9: False positive breakdown of Viola Jones eland classifier on the Simpson dataset.

metric, *precision*, which is the percentage of true positive detections out of the total number of detections. From Figure 6.10, we can see that the precision on the Simpson dataset was 20.52% compared to 47.16% during cross-validation. The precision during cross-validation was higher because the positive and negative test set sizes were kept equal, however, as we can see in the Simpson dataset, there is a significantly lower proportion of positive images compared to the negative images.

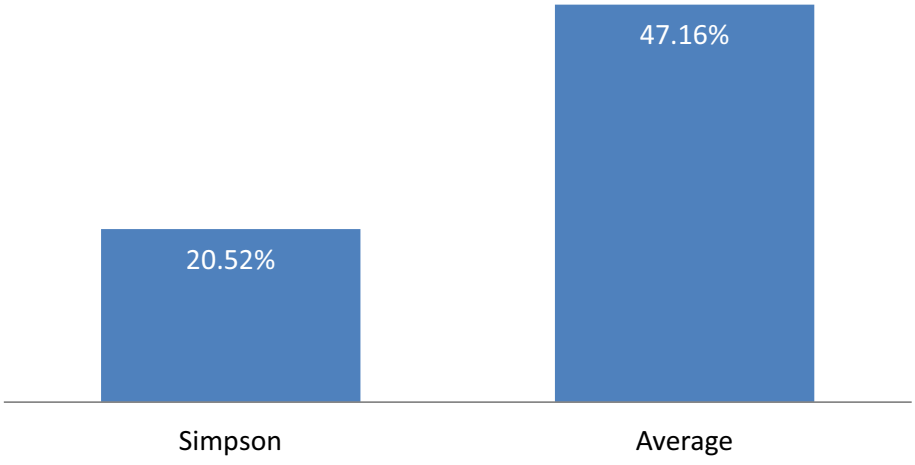
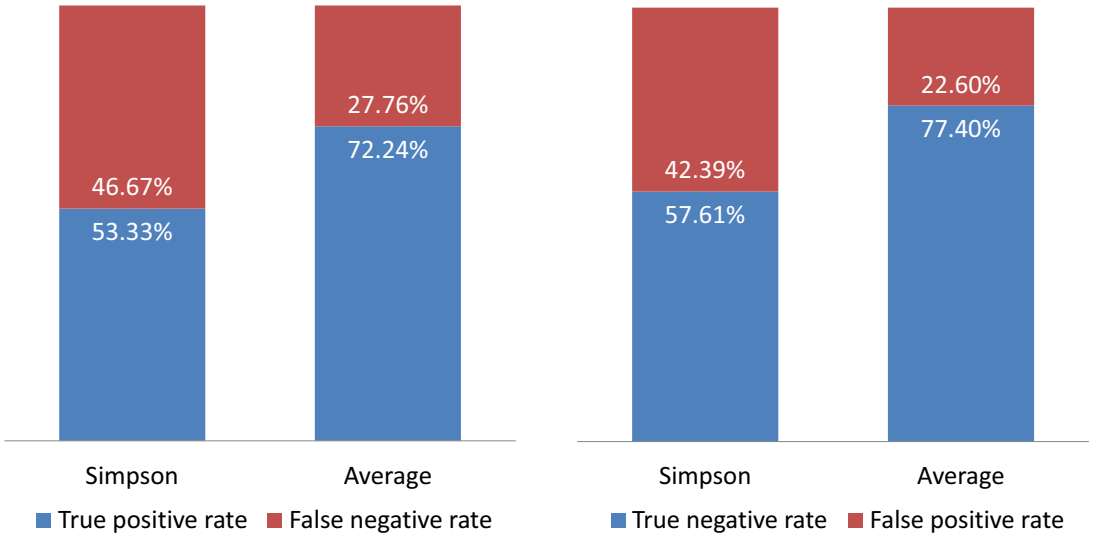


Figure 6.10: Viola Jones eland classifier precision on the Simpson dataset compared to its average precision in cross-validation.

6.4.2 Support vector machine performance

The support vector machine classifier was passed the 514 detections from the Viola Jones classifier, 120 of which were true positives while the other 394 were false positives. From Graph 6.11, we can see that the true positive rate of the support vector machine classifier was 53.33%, significantly lower than the 72.24% from the results obtained during cross-validation. Similarly the true negative rate was 57.61%, also significantly lower than the 77.4% obtained during cross-validation. The true positive rate was most likely lower because the sub-windows passed from the Viola Jones classifier did not always perfectly enclose the eland like they were in the training sets. For example, sometimes the head of the eland would be just out of frame with the other side of the image padded with a little extra blank rock space. Thus, the pixel intensities in these images would be slightly offset and when projected in the higher dimensional space, it would fall on the wrong side of the separating hyperplane. As for the reduction in true negative rate, it was most likely because there were variations in the Simpson dataset that were not present in our training sets. This resulted in the classifier only performing slightly better than guessing since those variations in the pixel intensities in hyperspace could fall on either side of the separating hyperplane.



(a) Positive test set performance

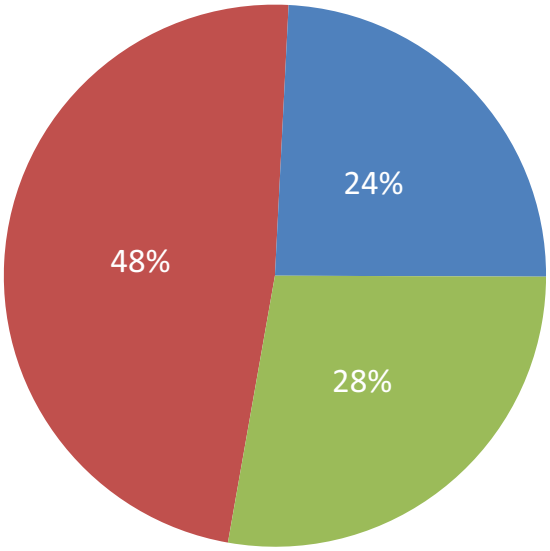
(b) Negative test set performance

Figure 6.11: Support vector machine eland classifier performance on the Simpson dataset compared to its average performance in cross-validation.

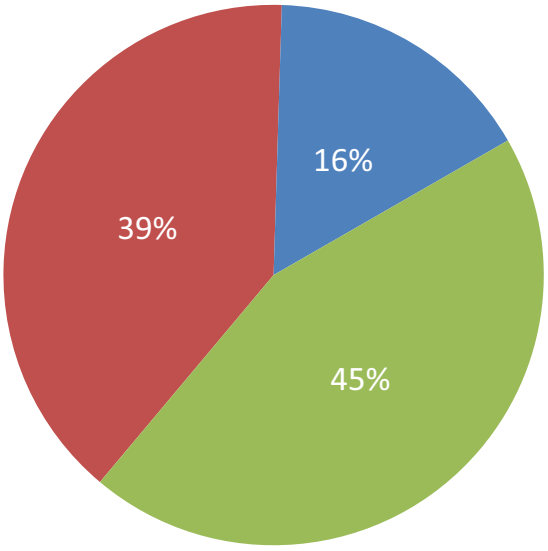
6.4.3 Overall performance

In this section, we look at the overall performance of our two-stage classifier by analysing the number of eland that passed through both the Viola Jones and support vector machine classifier. From Figure 6.12, we can see that 28% of the eland, which equates to 43 eland, in the Simpson dataset successfully passed through both classifiers which compares to 45% during cross-validation. As explained in the preceding sections, the lower true positive

rate was caused by the presence of outlier images with a large number of eland and the imperfection of the sub-windows passed from the Viola Jones classifier. These problems were further compounded by the deterioration and the intra-class variation in the art.



(a) Simpson dataset. Legend: Eland rejected by Viola Jones classifier. Eland accepted by Viola Jones classifier but rejected by SVM. Eland accepted by Viola Jones classifier and SVM.



(b) Average. Legend: Eland rejected by Viola Jones classifier. Eland accepted by Viola Jones classifier but rejected by SVM. Eland accepted by Viola Jones classifier and SVM.

Figure 6.12: Breakdown of true positives on the Simpson dataset compared to the average during cross-validation.

6.5 Conclusion

In conclusion, our first stage Viola Jones classifier had an average true positive rate, between the four classifiers for eland, elephants, humans, and rhebuck, of almost 60% but at the expense of just under 1 false positive per image. We found that the true positive rate was 60% due to the deterioration and variation in the art. We also found that in the case of the elephant classifier, the Viola Jones framework finds engravings particularly challenging to identify. As for the false positives per image, we concluded that rough and ragged rock surfaces cast small shadows across paintings and blank rock surfaces which get confused as features of the object and similar looking objects, especially quadrupeds, are sometimes confused for each other. A closer analysis of the training set sizes revealed that as the positive sample set size increases, the true positive rate begins to taper off after approximately 800 samples but there is a significant drop in the false positive rate per image when the positive sample set size increases to over 1400. The support vector machine, which was passed the sub-windows detected by the Viola Jones classifier, had an average true positive rate of 72% and a true negative rate of 77% which had the effect of eliminating over three-quarters of the false positives from the Viola Jones classifier at the expense of just over a quarter of the true positives eliminated as well. Lastly, we

found that the performance of the eland classifier on an unlabelled Simpson sub-dataset of the SARADA, the performances of the classifiers were not as expected. The Viola Jones classifier had a similar true positive and false positive rate but the significantly higher proportion of negative images in the set led to the false positives dominating the total number of detections. Only 21% of the detections ended up relevant. We hoped that the support vector machine would mitigate this and remove most of the false positives at a low cost to the true positives but it ended up only performing slightly better than guessing with a true positive and false positive rate of 53% and 58% respectively. We concluded that there were unique variations in the Simpson dataset that were not encoded in our training sets and the imperfection of the true positive sub-windows passed into the support vector machine caused problems which were further compounded by the fundamental characteristics of the rock art, deterioration and variation.

In the next chapter, we conclude this dissertation by answering the research questions, discuss the contribution of this work, and outline potential future work.

Chapter 7

Conclusion and Future Work

Despite advances in image processing and machine learning, very little attention has been paid to the field of rock art object detection. There are several challenges posed by this domain such as the intra-class variation of objects, partial occlusion, arbitrary scale and rotation, and deterioration. Previous attempts sidestepped most of these challenges and instead just focused on the recognition of digitised black and white reliefs, whether it be from the end user or from teams of specialised rock art researchers.

In this dissertation, we have demonstrated the application of a combination of two supervised learning methods for object detection and recognition of rock art in their original form. Specifically, we have used the Viola Jones object detection framework to search through an image at all scales to detect and initially classify a sub-window which was then passed on to an SVM to provide a final classification. We implemented our experiments as a five-fold cross validation process to assess the performance of these classifiers. With this completed, we are in a position to answer the research questions set out in Section 5.5.

1. Can the Viola Jones object detection framework trained with Haar-like features detect and recognise multiple rock art objects in full raw photographs?

   The Viola Jones object detection framework trained with Haar-like features could detect and correctly classify multiple instances of an object in a full raw photograph. It located and classified 59.44% of the objects which were found to usually be of objects that were still relatively intact compared to the ones that were missed which were severely deteriorated in the form of flaking or fading. The Viola Jones classifier also detected 0.96 false positives per image which were usually of similar looking objects are regions with jagged rock surfaces that cast small shadows that were mistaken for features. In practice, however, we found that there were relatively few instances of an object compared to the number of images in a dataset which resulted in our classifiers having a much lower precision.

2. Can a secondary classifier, such as an SVM, which is passed the sub-windows detected by the Viola Jones classifier, improve the classification performance?

   The SVM was trained and tested using pixel intensity vectors of sub-windows containing exactly one full object (for the positive sample set) or none at all (for the negative sample set). It performed reasonably well in testing with a true positive

rate of 72.24% and a true negative rate of 77.4%. Therefore, we predicted that over three-quarters of the false positives from the Viola Jones classifier would be correctly rejected but at the expense of about a quarter of the true positives thrown out as well which would result in a higher precision of the left over detections. However, when testing the Viola Jones and SVM combination on an unlabelled portion of the SARADA, the performance dropped to a true positive and negative rate of 53.33% and 57.61% respectively. We attributed this to the imperfect enclosure of the objects in the sub-windows passed from the Viola Jones classifier and different variations in the unlabelled dataset not encoded in our training sets. Therefore, to answer the question, in our controlled experiments the SVM did improve the overall performance but in practice it did not perform as well.

Overall, we find our results to be promising considering all the difficulties posed in this domain. We believe that we have laid the ground work upon which future studies could be carried out in the detection and recognition of rock art paintings. Lines of work worth investigating are partial shape matching techniques for the recognition of deteriorated paintings and rotational invariance to paintings drawn sideways or upside down, an occurrence that is fairly common in southern African rock art. Also worth investigating is an automatic rock art segmentation system specifically tailored for paintings. It has already been shown to be possible for engravings and if so for paintings, the removal of extraneous background noise should simplify the classification of an object.

Appendix A

Viola Jones Cross Validation Results

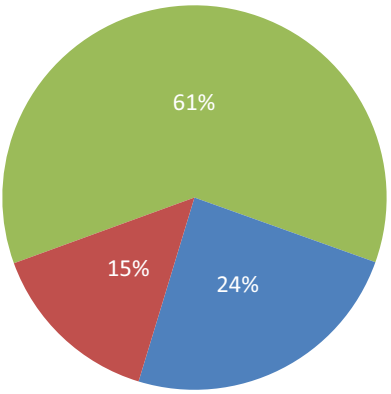
Eland			Elephants		
Set	TP (%)	FP/Img	Set	TP (%)	FP/Img
1	64.44	0.59	1	53.33	1.13
2	62.56	0.65	2	54.35	1.06
3	51.91	0.68	3	36.59	0.68
4	62.95	0.55	4	45.65	1.15
5	62.75	0.67	5	43.48	0.80
Avg	60.92	0.63	Avg	46.68	0.96

Humans			Rhebuck		
Set	TP (%)	FP/Img	Set	TP (%)	FP/Img
1	62.84	1.16	1	64.44	0.96
2	68.03	1.27	2	70.45	1.14
3	65.61	1.03	3	65.31	1.12
4	65.52	1.29	4	67.39	1.01
5	55.24	0.89	5	65.91	1.33
Avg	63.45	1.13	Avg	66.70	1.11

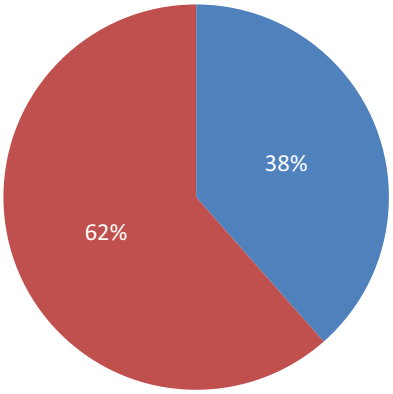
Table A.1: Viola Jones cross-validation results.

Appendix B

Viola Jones False Positive Graphs

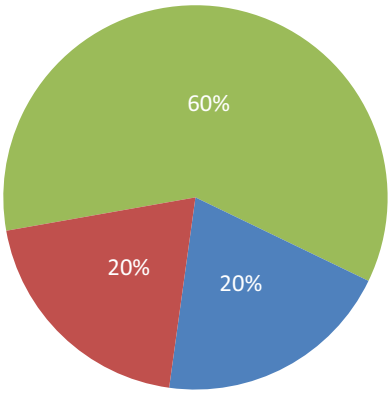


(a) Images with 0 FP's. Images with 1 FP. Images with more than 1 FP.

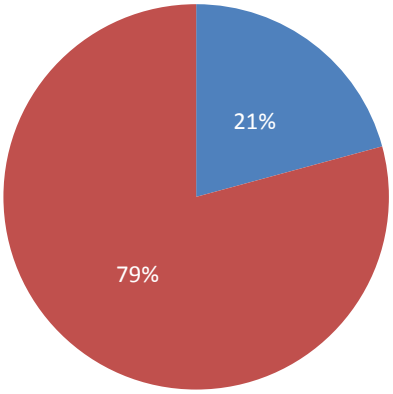


(b) FP's from images with 1 FP. FP's from images with more than 1 FP.

Figure B.1: Eland false positive analysis.

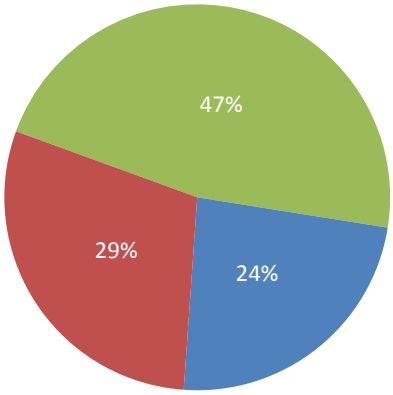


(a) Images with 0 FP's. Images with 1 FP. Images with more than 1 FP.

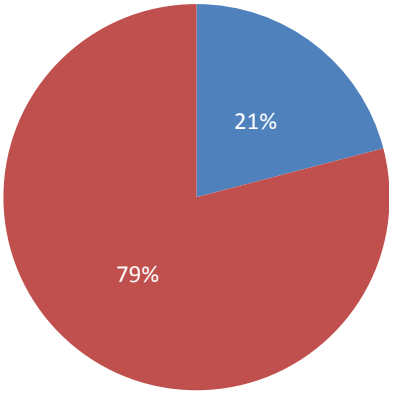


(b) FP's from images with 1 FP. FP's from images with more than 1 FP.

Figure B.2: Elephant false positive analysis.

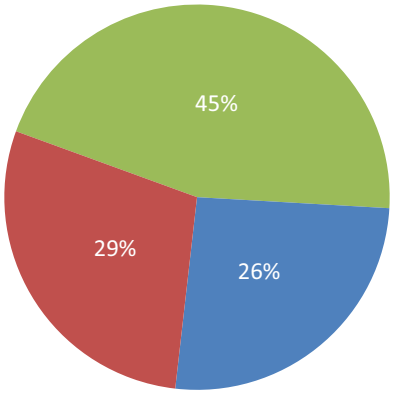


(a) Images with 0 FP's. Images with 1 FP. Images with more than 1 FP.

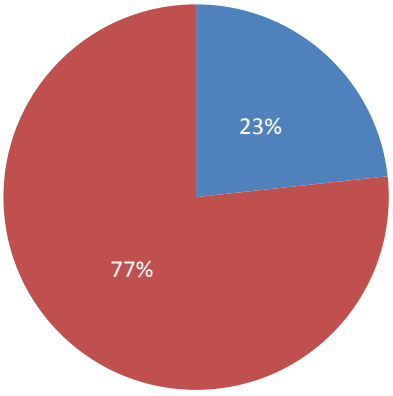


(b) FP's from images with 1 FP. FP's from images with more than 1 FP.

Figure B.3: Humans false positive analysis.



(a) Images with 0 FP's. Images with 1 FP. Images with more than 1 FP.



(b) FP's from images with 1 FP. FP's from images with more than 1 FP.

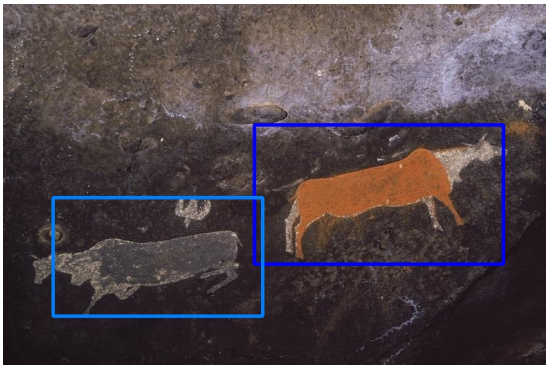
Figure B.4: Rhebuck false positive analysis.

Appendix C

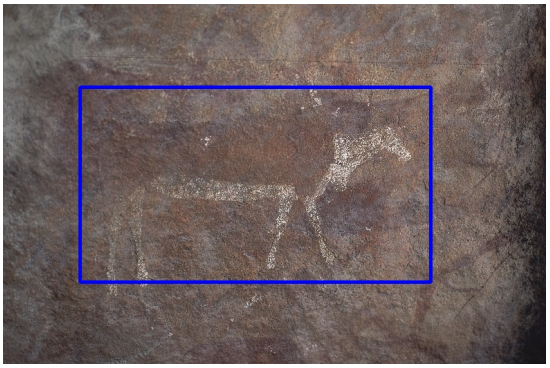
Viola Jones True Positives



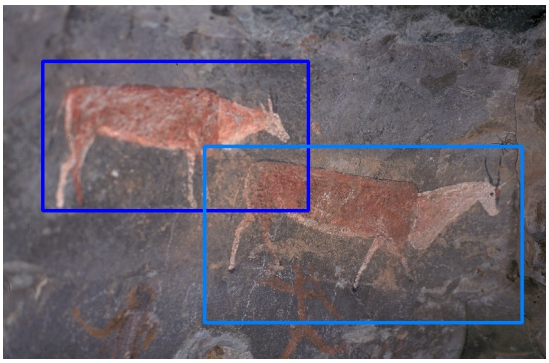
(a) NMSA-PJV-01-163HC



(b) RARI-LEE-RSA-MAT1-30

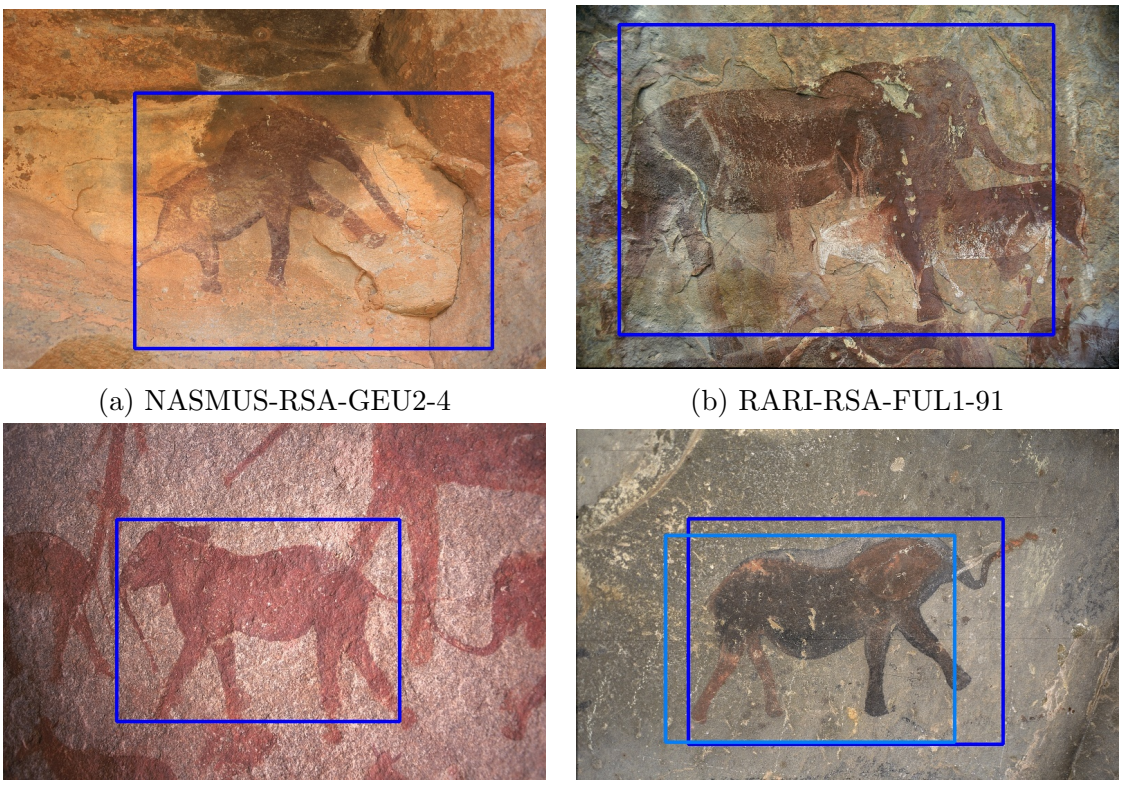


(c) RARI-RSA-GQA11-38



(d) RARI-RSA-FST3-1

Figure C.1: The true positive detections of eland did seem to be of images that were clear and of high contrast such as in C.1a, C.1b, and C.1d. A handful of detections were also of images that were highly deteriorated such as in C.1c. We can also see examples of multiple detections in C.1b and C.1d with the former containing eland facing opposite directions.
Images reproduced with permission from the Natal Museum (Department of Archaeology & Anthropology), Pietermaritzburg, South Africa (C.1a), and the Rock Art Research Institute, University of the Witwatersrand, South Africa (C.1b, C.1c, C.1d). Retrieved 28 October 2014, from SARADA, www.sarada.co.za.



(a) NASMUS-RSA-GEU2-4

(b) RARI-RSA-FUL1-91

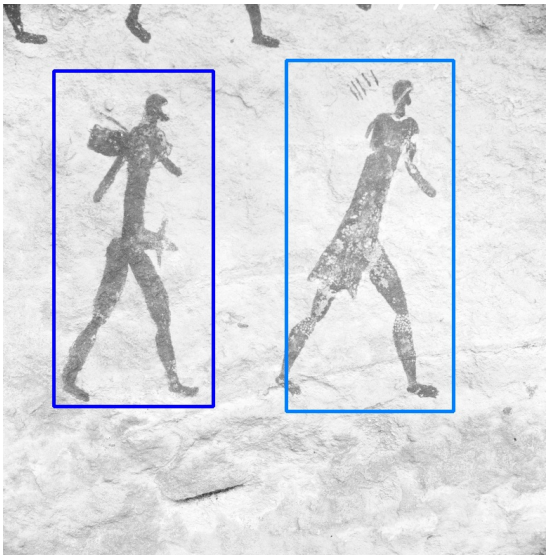
(c) RARI-RSA-LOW3-30

(d) RSA-EMB1-31

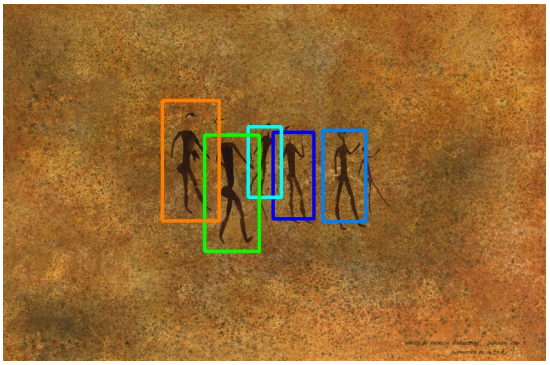
Figure C.2: The true positive detections of elephant were usually monochrome paintings that were still relatively intact.
Images reproduced with permission from the National Museum, Bloemfontein, South Africa (C.2a) and the Rock Art Research Institute, University of the Witwatersrand, South Africa (C.2b, C.2c, C.2d). Retrieved 28 October 2014, from SARADA, www.sarada.co.za.



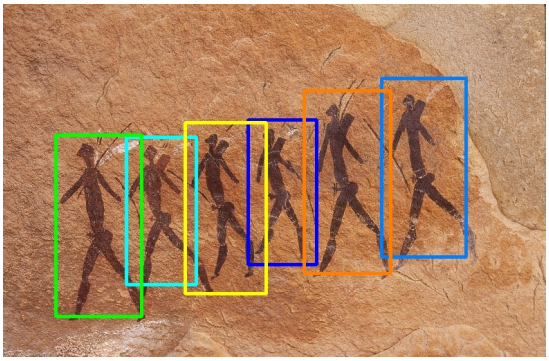
(a) NASMUS-RSA-HOF1-16



(b) NMSA-PCA-01-66H

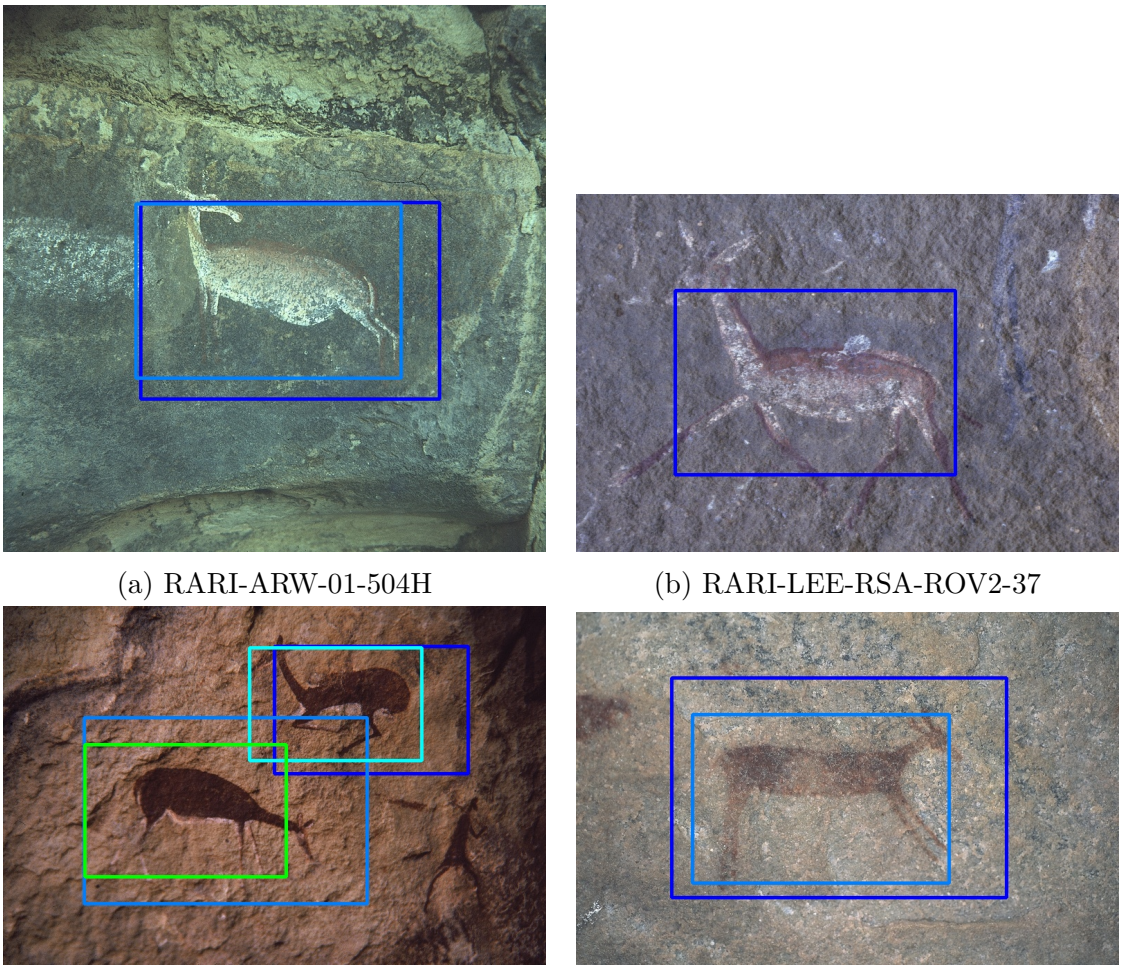


(c) NMSA-PJV-01-73HC



(d) RSA-HUT1-2

Figure C.3: The true positive detections of humans standing or walking were clear and largely intact. The depictions of these vary in their dress and the equipment they carry. For example, headgear and waist bands can be seen in C.3a while the a cloak can be seen in C.3b. Bows and quivers can also be seen in C.3d.
Images reproduced with permission from the National Museum, Bloemfontein, South Africa (C.3a), the Natal Museum (Department of Archaeology & Anthropology), Pietermaritzburg, South Africa (C.3b, C.3c), and the Rock Art Research Institute, University of the Witwatersrand, South Africa (C.3d). Retrieved 28 October 2014, from SARADA, www.sarada.co.za.



(a) RARI-ARW-01-504H

(b) RARI-LEE-RSA-ROV2-37

(c) RARI-LEE-RSA-XEN1-7

(d) RARI-RSA-GQA5-71

Figure C.4: Examples of true positive detections of rhebuck showing shaded polychrome (C.4a, C.4b), bichrome (C.4c), and monochrome (C.4d) paintings. Figures C.4a, C.4c, and C.4d show multiple detections of the same rhebuck that were considered one detection.
Images reproduced with permission from the Rock Art Research Institute, University of the Witwatersrand, South Africa. Retrieved 28 October 2014, from SARADA, www.sarada.co.za.

Appendix D

Viola Jones False Negatives



(a) IZI-HT-01-51HC

(b) JDC-02-164

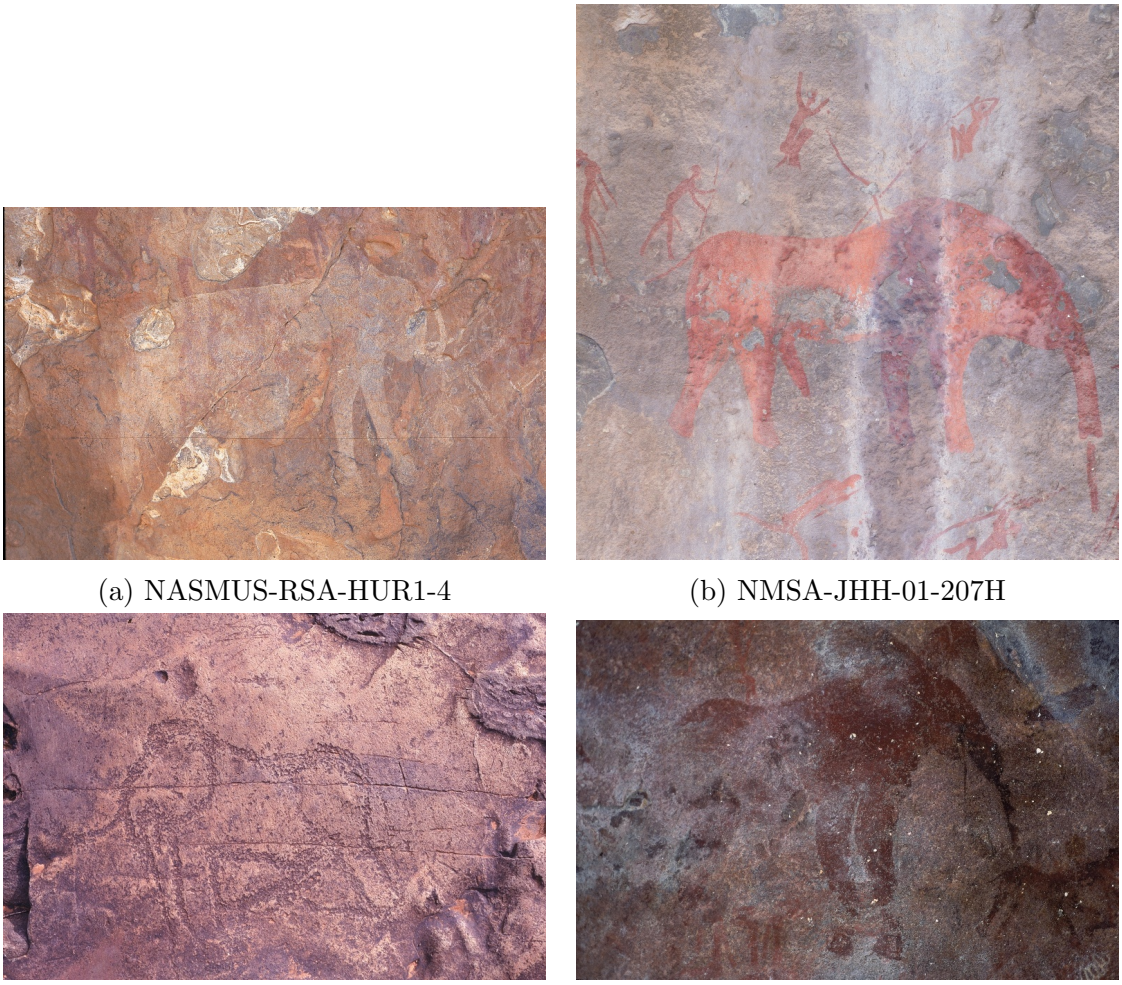


(c) RARI-RSA-RUT1-34



(d) RSA-NQA1-4

Figure D.1: The eland false negatives here are the same colour as their background (D.1a), have a low contrast with their background in grayscale (reds and dark backgrounds have similar shades) (D.1b, D.1d), or were faded and on rough rock surfaces (D.1c).
Images reproduced with permission from the Iziko Museums of Cape Town Social History Collections, South Africa (D.1a), Janette Deacon Archive, South Africa (D.1b), and Rock Art Research Institute, University of the Witwatersrand, South Africa (D.1c, D.1d). Retrieved 28 October 2014, from SARADA, www.sarada.co.za.



(a) NASMUS-RSA-HUR1-4

(b) NMSA-JHH-01-207H

(c) RARI-RSA-STY1-6

(d) ZIM-LSR32-12

Figure D.2: The elephant false negatives were either extremely faded (D.2a), partially deteriorated (D.2b, D.2d), or engravings (D.2c).
Images reproduced with permission from the National Museum, Bloemfontein, South Africa (D.2a), the Natal Museum (Department of Archaeology & Anthropology), Pietermaritzburg, South Africa (D.2b), and the Rock Art Research Institute, University of the Witwatersrand, South Africa (D.2c, D.2d). Retrieved 28 October 2014, from SARADA, www.sarada.co.za.



(a) NASMUS-RSA-ANG1-32



(b) RARI-RSA-GQA1-59

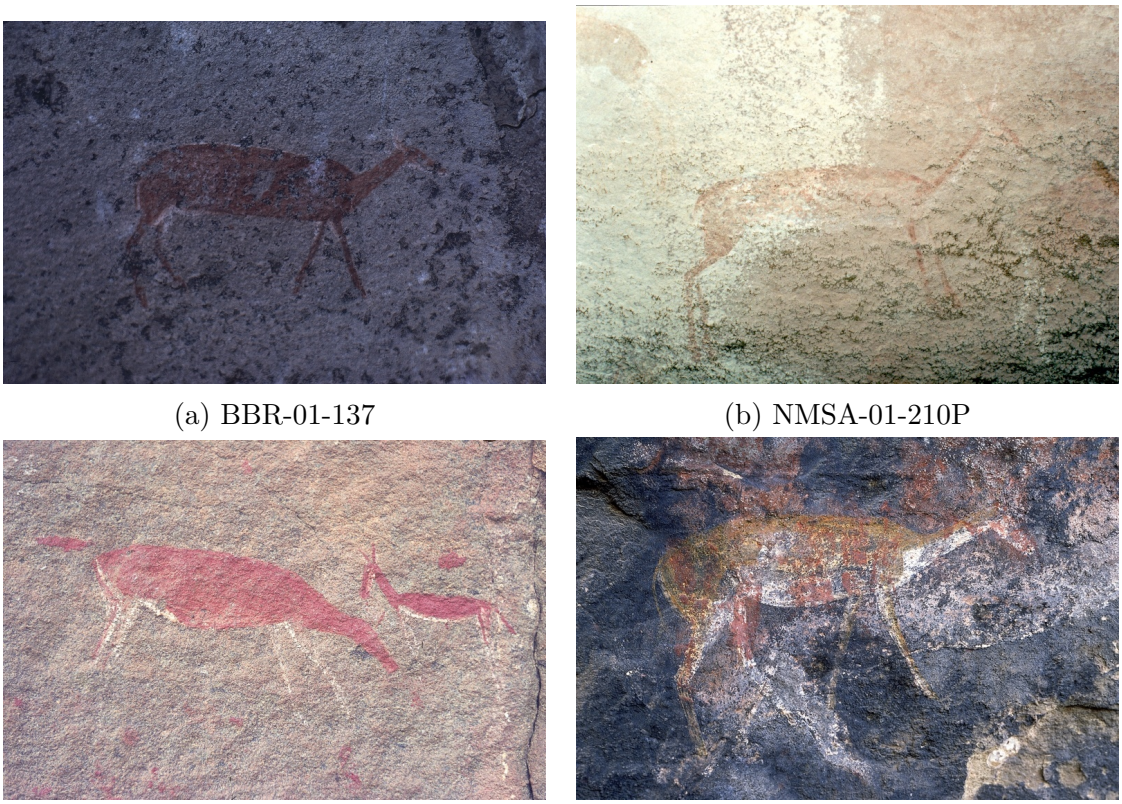


(c) RSA-MEL6-44



(d) RSA-MOI1-4

Figure D.3: The human false negatives were usually painted on rough rock surfaces as all these examples show, partially deteriorated with the arms missing as in D.3a, or were standing in a slight bent over posture as in D.3c and D.3d.
Images reproduced with permission from the National Museum, Bloemfontein, South Africa (D.3a), and the Rock Art Research Institute, University of the Witwatersrand, South Africa (D.3b, D.3c, D.3d). Retrieved 28 October 2014, from SARADA, www.sarada.co.za.



(a) BBR-01-137

(b) NMSA-01-210P

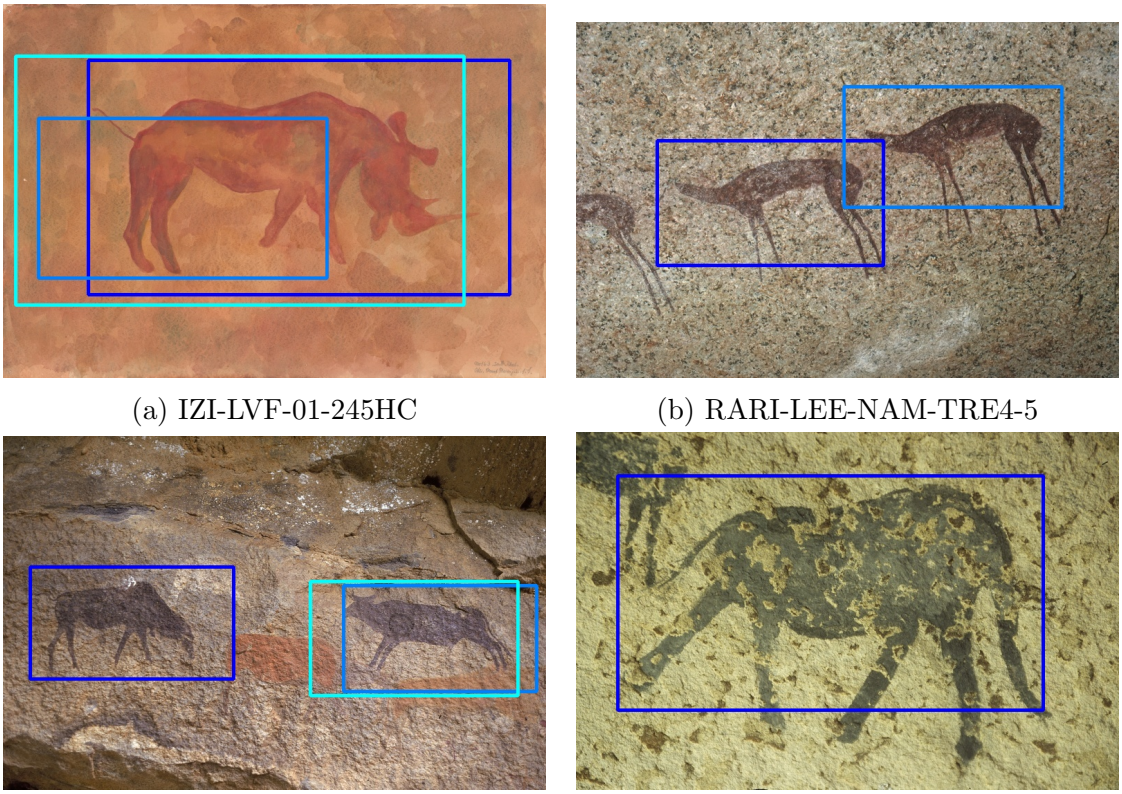
(c) RARI-LEE-RSA-DRY1-22

(d) RARI-LEE-RSA-WEG3-20

Figure D.4: The rhebuck false negatives were also either faded as in D.4b or on a rough and variably coloured rock surface as in D.4d. However, in D.4a and D.4c, the artwork seems to be relatively intact and on a smooth surface. Perhaps the classifier was simply under trained and required more training samples.
Images reproduced with permission from the Bill Barnes Archive, South Africa (D.4a), Natal Museum (Department of Archaeology & Anthropology), Pietermaritzburg, South Africa (D.4b), and Rock Art Research Institute, University of the Witwatersrand, South Africa (D.4c, D.4d). Retrieved 28 October 2014, from SARADA, www.sarada.co.za.

Appendix E

Viola Jones False Positives



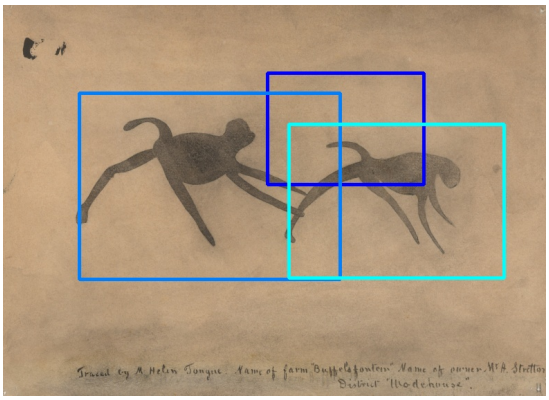
(a) IZI-LVF-01-245HC

(b) RARI-LEE-NAM-TRE4-5

(c) RARI-LEE-RSA-MRE1-8

(d) RARI-LEE-RSA-POK1-19

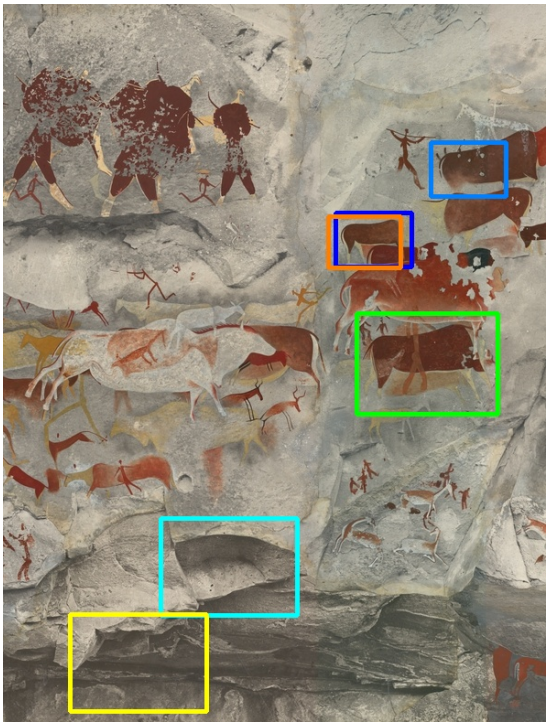
Figure E.1: Examples of false positives from the eland classifier where it mistakenly identified a rhino (E.1a), rhebuck (E.1b), cattle (E.1c), and an elephant (E.1d).
Images reproduced with permission from the Iziko Museums of Cape Town Social History Collections, South Africa (E.1a), and the Rock Art Research Institute, University of the Witwatersrand, South Africa (E.1b, E.1c, E.1d). Retrieved 28 October 2014, from SARADA, www.sarada.co.za.



(a) IZI-HT-01-101HC



(b) RARI-LEE-RSA-BUF1-12

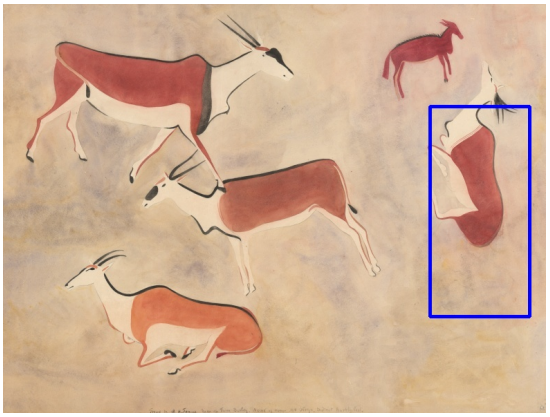


(c) RARI-HPC-01-23HC

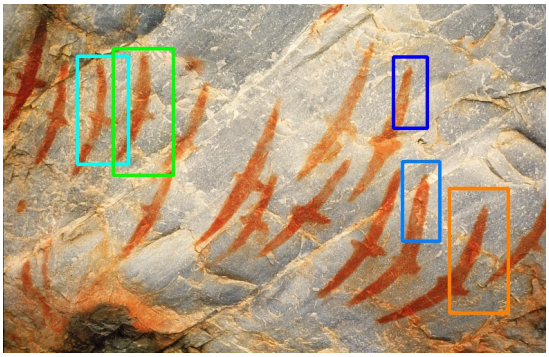


(d) RARI-LEE-RSA-GAB1-16

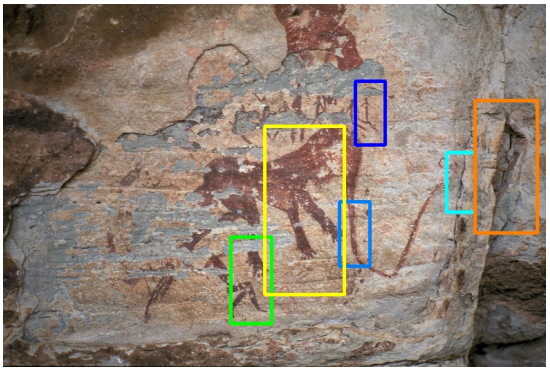
Figure E.2: Examples of false positives from the elephant classifier where it mistakenly identified some baboons (E.2a), a cow (E.2b), eland and blank rocks (E.2c), and two humans (E.2d).
Images reproduced with permission from the Iziko Museums of Cape Town Social History Collections, South Africa (E.2a), and the Rock Art Research Institute, University of the Witwatersrand, South Africa (E.2b, E.2c, E.2d). Retrieved 28 October 2014, from SARADA, www.sarada.co.za.



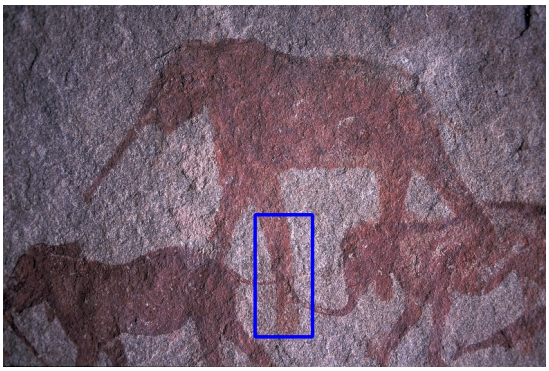
(a) IZI-HT-01-14HC



(b) RARI-LEE-RSA-BEK1-2

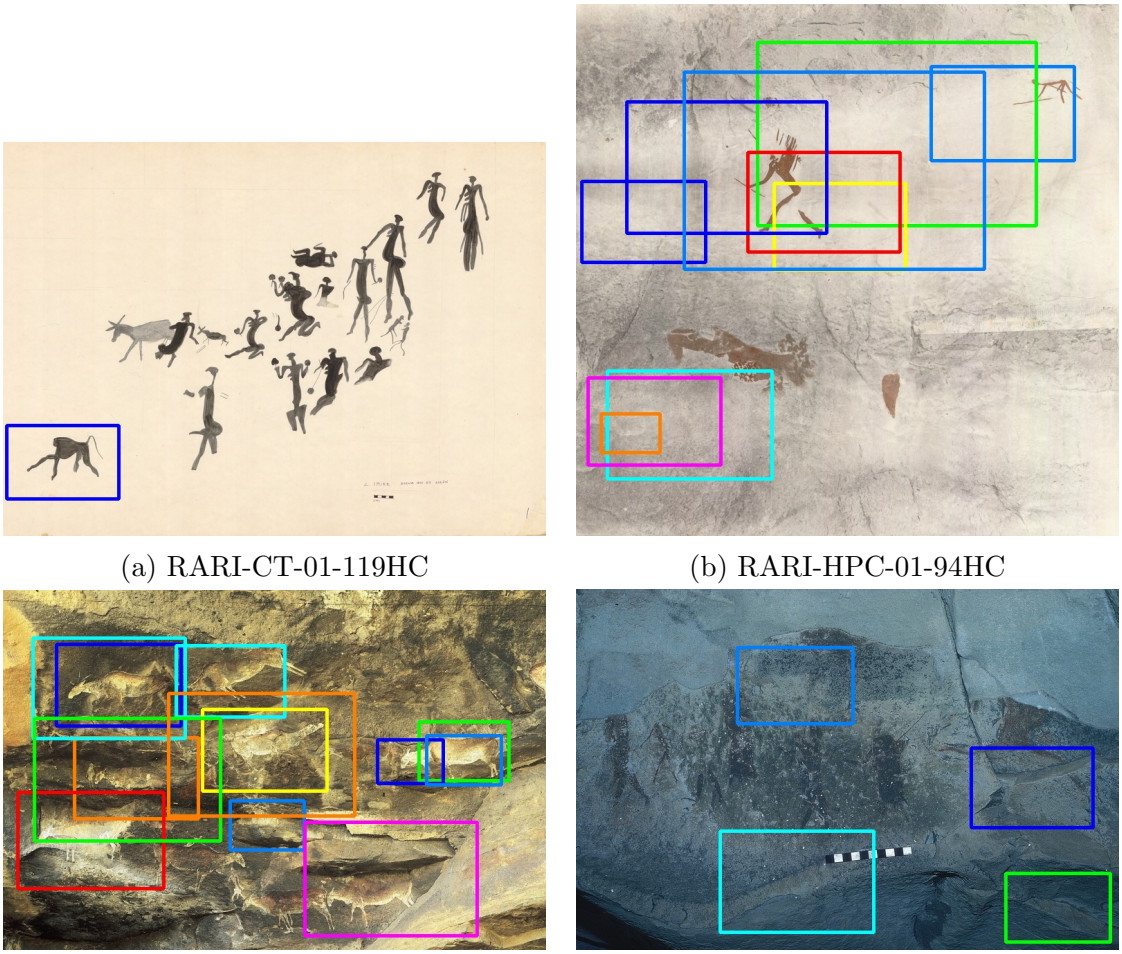


(c) RARI-RSA-BAR1-35



(d) RARI-RSA-LOW3-5

Figure E.3: Examples of the human false positives where it seems to have mistakenly identified other vertically shaped objects such as an eland facing upwards (E.3a), sideways birds with their wings outstretched (E.3b), the front of an elephant (E.3c), and the leg of another elephant (E.3d).
Images reproduced with permission from the Iziko Museums of Cape Town Social History Collections, South Africa (E.3a), and the Rock Art Research Institute, University of the Witwatersrand, South Africa (E.3b, E.3c, E.3d). Retrieved 28 October 2014, from SARADA, www.sarada.co.za.



(a) RARI-CT-01-119HC

(b) RARI-HPC-01-94HC

(c) RSA-BAE2-3

(d) RSA-BNO1-25

Figure E.4: Examples of rhebuck false positives where the classifier mistakenly identified a baboon in E.4a. The other three images are characterised by multiple false detections on a blank cave wall in E.4b and E.4d and on deteriorated eland in E.4c.
Images reproduced with permission from the Rock Art Research Institute, University of the Witwatersrand, South Africa (E.4a, E.4c, E.4c) and the Natal Museum (Department of Archaeology & Anthropology), Pietermaritzburg, South Africa (E.4d). Retrieved 28 October 2014, from SARADA, www.sarada.co.za.

Appendix F

Viola Jones True Negatives



(a) RARI-RSA-LOW3-5



(b) RSA-CSF11-36



(c) RSA-LAB1-113

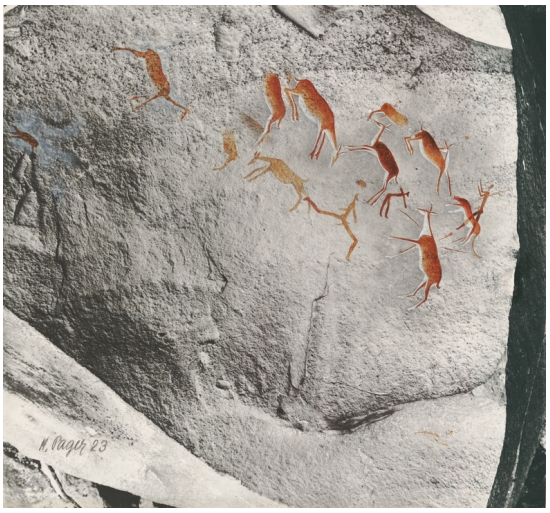


(d) SMT-01-3095

Figure F.1: Examples of true negatives from the eland classifier where it correctly rejected other quadrupeds such as elephants (F.1a), rain animals (mythical San creatures) (F.1b, F.1c), and baboons (F.1d).
Images reproduced with permission from the Rock Art Research Institute, University of the Witwatersrand, South Africa (F.1a, F.1c), Natal Museum (Department of Archaeology & Anthropology), Pietermaritzburg, South Africa (F.1b), and the Lucas Smits Archive, Netherlands (F.1d). Retrieved 28 October 2014, from SARADA, www.sarada.co.za.



(a) JHH-01-706H



(b) RARI-HPC-01-34HC



(c) RARI-RSA-FET3-82



(d) RARI-RSA-ORG1-32

Figure F.2: Examples of true negatives from the elephant classifier where it correctly rejected a horse (F.2a), rhebuck (F.2b), an eland (F.2c), and humans (F.2d).
Images reproduced with permission from the John Hone Archive, KwaZulu-Natal, South Africa (F.2a), and the Rock Art Research Institute, University of the Witwatersrand, South Africa (F.2b, F.2c, F.2d). Retrieved 28 October 2014, from SARADA, www.sarada.co.za.



(a) BBR-01-15



(b) JDC-RSA-BUT1-3



(c) RSA-LAB1-21D



(d) UOC-SRZ-01-2911P

Figure F.3: Examples of true negatives from the human classifier where it correctly rejected animals such as a rhebuck (F.3a), eland (F.3b), a lion (F.3c), and a rhino (F.3d). Images reproduced with permission from the Bill Barnes Archive, South Africa (F.3a), Janette Deacon Archive, South Africa (F.3b), Rock Art Research Institute, University of the Witwatersrand, South Africa (F.3c), and University of Cologne, Germany (F.3d). Retrieved 28 October 2014, from SARADA, www.sarada.co.za.



(a) JDC-RSA-KSD1-1

(b) JHH-01-133H

(c) RARI-LEE-RSA-BEL1-13

(d) RARI-LEE-RSA-VRE1-36

Figure F.4: Examples of true negatives from the rhebuck classifier where it correctly rejected a baboon (F.4a), a human (F.4b), a horse (F.4c), and more humans (F.4d). Images reproduced with permission from the Janette Deacon Archive, South Africa (F.4a), John Hone Archive, KwaZulu-Natal, South Africa (F.4b), and Rock Art Research Institute, University of the Witwatersrand, South Africa (F.4c, F.4d). Retrieved 28 October 2014, from SARADA, www.sarada.co.za.

Appendix G

Support Vector Machine Cross Validation Results

Eland			
Set	TP (%)	TN (%)	Average
1	69.64	69.33	69.49
2	72.40	73.68	73.04
3	78.08	69.33	73.71
4	70.75	72.00	71.38
5	75.38	72.00	73.69
Avg	73.25	71.27	72.26

Elephant			
Set	TP (%)	TN (%)	Average
1	78.05	72.00	75.03
2	61.90	89.80	75.85
3	77.50	74.00	75.75
4	76.74	68.00	72.37
5	61.90	66.00	63.95
Avg	71.22	73.96	72.59

Humans			
Set	TP (%)	TN (%)	Average
1	68.92	96.88	82.90
2	73.47	95.00	84.24
3	71.34	94.38	82.86
4	70.34	93.13	81.74
5	69.93	92.50	81.22
Avg	70.80	94.38	82.59

Rhebuck			
Set	TP (%)	TN (%)	Average
1	77.78	76.00	76.89
2	70.45	78.00	74.23
3	71.43	64.00	67.72
4	78.26	60.00	69.13
5	70.45	72.00	71.23
Avg	73.67	70.00	71.84

Table G.1: SVM cross-validation results.

Bibliography

[Adamek and O'Connor 2003] Tomasz Adamek and Noel O'Connor. Efficient contour-based shape representation and matching. In *Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, pages 138–143. ACM, 2003. 18

[Adamek and O'Connor 2004] Tomasz Adamek and Noel E O'Connor. A multiscale representation method for nonrigid shapes with a single closed contour. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(5):742–753, 2004. 18

[Attalla and Siy 2005] Emad Attalla and Pepe Siy. Robust shape similarity retrieval based on contour segmentation polygonal multiresolution and elastic matching. *Pattern Recognition*, 38(12):2229–2241, 2005. 18

[Aubert *et al.* 2014] M. Aubert, A. Brumm, M. Ramli, T. Sutikna, E.W. Saptomo, B. Hakim, M.J. Morwood, G.D. van den Bergh, L. Kinsley, and A. Dosseto. Pleistocene cave art from Sulawesi, Indonesia. *Nature*, 514(7521):223–227, 2014. 1

[Bagon *et al.* 2010] Shai Bagon, Ori Brostovski, Meirav Galun, and Michal Irani. Detecting and sketching the common. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 33–40. IEEE, 2010. 25

[Bai *et al.* 2007] Xiang Bai, Longin Jan Latecki, and Wen-Yu Liu. Skeleton pruning by contour partitioning with discrete curve evolution. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(3):449–462, 2007. 21

[Bay *et al.* 2006] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *Computer Vision–ECCV 2006*, pages 404–417. Springer, 2006. 25

[Belongie *et al.* 2002] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4):509–522, 2002. 17, 21, 22

[Boser *et al.* 1992] Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pages 144–152. ACM, 1992. 36

[Brown *et al.* 1999] Michael P.S. Brown, William Noble Grundy, David Lin, Nello Cristianini, Charles Sugnet, Manuel Ares, and David Haussler. Support vector machine

classification of microarray gene expression data. *University of California, Santa Cruz, Technical Report UCSC-CRL-99-09*, 1999. 36

[Burges 1998] Christopher J.C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998. 34

[Campbell 2002] Colin Campbell. Kernel methods: a survey of current techniques. *Neurocomputing*, 48(1):63–84, 2002. 36

[Canny 1986] John Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):679–698, 1986. 24

[Chang and Lin 2011] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm. 43

[Cortes and Vapnik 1995] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995. 2, 26, 34, 38

[Crow 1984] Franklin C. Crow. Summed-area tables for texture mapping. *ACM SIGGRAPH computer graphics*, 18(3):207–212, January 1984. 27

[Deufemia and Paolino 2013] Vincenzo Deufemia and Luca Paolino. Combining unsupervised clustering with a non-linear deformation model for efficient petroglyph recognition. In *Advances in Visual Computing*, pages 128–137. Springer, 2013. 17

[Deufemia and Paolino 2014] Vincenzo Deufemia and Luca Paolino. Segmentation and recognition of petroglyphs using generic fourier descriptors. In *Image and Signal Processing*, pages 487–494. Springer, 2014. 20

[Deufemia *et al.* 2012a] Vincenzo Deufemia, Luca Paolino, and Henry de Lumley. Petroglyph recognition using self-organizing maps and fuzzy visual language parsing. In *Tools with Artificial Intelligence (ICTAI), 2012 IEEE 24th International Conference*, volume 1, pages 852–859. IEEE, 2012. 16

[Deufemia *et al.* 2012b] Vincenzo Deufemia, Luca Paolino, Genoveffa Tortora, Antonella Traverso, Viviana Mascardi, Massimo Ancona, Maurizio Martelli, Nicoletta Bianchi, and Henry De Lumley. Investigative analysis across documents and drawings: visual analytics for archaeologists. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 539–546. ACM, 2012. 23

[Deufemia *et al.* 2014a] Vincenzo Deufemia, Viviana Mascardi, Luca Paolino, Giuseppe Polese, and Henry de Lumley. A volunteered geographic information system for collecting and rating petroglyph data. *Journal of Visual Languages & Computing*, 2014. 23

[Deufemia *et al.* 2014b] Vincenzo Deufemia, Valentina Indelli Pisano, Luca Paolino, and Paola de Roberto. A visual analytics system for supporting rock art knowledge discovery. In *Computational Science and Its Applications–ICCSA 2014*, pages 466–480. Springer, 2014. 23

[Díaz and Castro 2002] Diego Díaz and Damián Castro. Pattern recognition applied to rock art. *BAR International Series*, 1016:463–468, 2002. 15

[Dorkó and Schmid 2003] Gyuri Dorkó and Cordelia Schmid. Selection of scale-invariant parts for object class recognition. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 634–639. IEEE, 2003. 16

[Fletcher 1987] Roger Fletcher. *Practical Methods of Optimization*. John Wiley and Sons, Inc., 2nd edition, 1987. 36

[Fornés 2009] A Fornés. *Writer identification by a combination of graphical features in the framework of old handwritten music scores*. PhD thesis, Universitat Autònoma de Barcelona, 2009. 18

[Freund and Schapire 1995] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Computational learning theory*, pages 23–37. Springer, 1995. 30

[Gdalyahu and Weinshall 1999] Yoram Gdalyahu and Daphna Weinshall. Flexible syntactic matching of curves and its application to automatic hierarchical classification of silhouettes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(12):1312–1328, 1999. 18

[Girosi *et al.* 1995] Federico Girosi, Michael Jones, and Tomaso Poggio. Regularization theory and neural networks architectures. *Neural computation*, 7(2):219–269, 1995. 35

[Haar 1910] Alfred Haar. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69(3):331–371, 1910. 27

[Hall *et al.* 2007] Kevin Hall, Ian Meiklejohn, and Joselito Arocena. The thermal responses of rock art pigments: Implications for rock art weathering in southern Africa. *Geomorphology*, 91(1):132–145, 2007. 12

[Haralick *et al.* 1973] Robert M Haralick, Karthikeyan Shanmugam, and Its' Hak Dinstein. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, SMC-3(6):610–621, 1973. 22

[Haykin 1998] Simon Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition, 1998. 37

[Hœrlé and Salomon 2004] Stéphane Hœrlé and Andrew Salomon. Microclimatic data and rock art conservation at Game Pass shelter in the Kamberg Nature Reserve, KwaZulu-Natal. *South African Journal of Science*, 100(7-8):340–341, 2004. 12

[Hœrlé 2005] S Hœrlé. A preliminary study of the weathering activity at the rock art site of Game Pass Shelter (KwaZulu-Natal, South Africa) in relation to its conservation. *South African Journal of Geology*, 108(2):297–308, 2005. 12

[Hœrlé 2006] Stéphane Hœrlé. Rock temperatures as an indicator of weathering processes affecting rock art. *Earth Surface Processes and Landforms*, 31(3):383–389, 2006. 12

[Hsu *et al.* 2003] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. *A practical guide to support vector classification*, 2003. 38

[Intel *et al.* 2014] Intel, Willow Garage, and Itseez. *OpenCV*, 2014. Software available at http://opencv.org/. 43

[Jalba *et al.* 2005] Andrei C Jalba, Michael HF Wilkinson, Jos BTM Roerdink, Micha M Bayer, and Stephen Juggins. Automatic diatom identification using contour analysis by morphological curvature scale spaces. *Machine Vision and Applications*, 16(4):217–228, 2005. 18

[Jones and Viola 2003] Michael Jones and Paul Viola. Fast multi-view face detection. *Mitsubishi Electric Research Lab TR-20003-96*, 3:14, 2003. 42

[Keogh and Ratanamahatana 2005] Eamonn Keogh and Chotirat Ann Ratanamahatana. Exact indexing of dynamic time warping. *Knowledge and information systems*, 7(3):358–386, 2005. 18

[Keysers *et al.* 2004] Daniel Keysers, Christian Gollan, and Hermann Ney. Local context in non-linear deformation models for handwritten character recognition. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pages 511–514. IEEE, 2004. 17

[Khosravi and Kabir 2007] Hossein Khosravi and Ehsanollah Kabir. Introducing a very large dataset of handwritten Farsi digits and a study on their varieties. *Pattern Recognition Letters*, 28(10):1133–1141, 2007. 18

[Kroon and Omlin 2004] Steve Kroon and Christian W. Omlin. Getting to grips with support vector machines: Theory. *South African Statistical Journal*, 282:159–172, 2004. 34

[Kroon 2003] Steve R. Kroon. *Support Vector Machines, Generalization Bounds and Transduction*. Master's thesis, Department of Computer Science, University of Stellenbosch, 2003. 37

[Lewis-Williams *et al.* 1988] J. D. Lewis-Williams, T.A. Dowson, Paul G. Bahn, H.-G. Bandi, Robert G. Bednarik, John Clegg, Mario Consens, Whitney Davis, Brigitte Delluc, Gilles Delluc, Paul Faulstich, John Halverson, Robert Layton, Colin Martindale, Vil Mirimanov, Christy G. Turner II, Joan M. Vastokas, Michael Winkelman, and Alison Wylie. The signs of all times: Entoptic phenomena in upper palaeolithic art. *Current Anthropology*, 29(2):pp. 201–245, 1988. 3

[Lewis-Williams 1990] J. D. Lewis-Williams. *Discovering Southern African Rock Art*. David Philip Publishers (Pty) Ltd, 208 Werdmuller Centre, Claremont, 7700, South Africa, 1990. 3, 6, 9, 12

[Lienhart *et al.* 2003] Rainer Lienhart, Er Kuranov, and Vadim Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In *In DAGM 25th Pattern Recognition Symposium*, pages 297–304, 2003. 29, 30, 46

[Ling and Jacobs 2007] Haibin Ling and David W Jacobs. Shape classification using the inner-distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(2):286–299, 2007. 21

[Lowe 1999] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999. 25

[Lowe 2004] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004. 22, 25

[Mascardi *et al.* 2014] Viviana Mascardi, Daniela Briola, Angela Locoro, Daniele Grignani, Vincenzo Deufemia, Luca Paolino, Nicoletta Bianchi, Henry de Lumley, Damiano Malafronte, and Alessandro Ricciarelli. A holonic multi-agent system for sketch, image and text interpretation in the rock art domain. *International Journal of Innovative Computing, Information and Control*, 2014. 21

[Niels *et al.* 2008] R.M.J. Niels, D.J.M. Willems, and L.G. Vuurpijl. The NicIcon database of handwritten icons. In *Proceedings of the 11th International Conference on the Frontiers of Handwriting Recognition*, pages 296–301. Montreal, Canada: Concordia University, 2008. 18

[Ojala *et al.* 2002] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, 2002. 22

[Oren *et al.* 1997] Michael Oren, Constantine Papageorgiou, Pawan Sinha, Edgar Osuna, and Tomaso Poggio. Pedestrian detection using wavelet templates. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 193–199. IEEE, 1997. 27

[Papageorgiou *et al.* 1998] Constantine P Papageorgiou, Michael Oren, and Tomaso Poggio. A general framework for object detection. In *Computer Vision, 1998. Sixth International Conference on*, pages 555–562. IEEE, 1998. 27

[Perona and Malik 1990] Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(7):629–639, 1990. 24

[Prewitt 1970] Judith M.S. Prewitt. Object enhancement and extraction. *Picture processing and Psychopictorics*, 10(1):15–19, 1970. 24

[Ratanamahatana and Keogh 2005] Chotirat Ann Ratanamahatana and Eamonn Keogh. Three myths about dynamic time warping data mining. In *Proceedings of SIAM International Conference on Data Mining (SDM05)*, pages 506–510. SIAM, 2005. 18

[Roberts 1963] Lawrence Gilman Roberts. *Machine Perception of Three-Dimensional Solids*. PhD thesis, Massachusetts Institute of Technology, 1963. 24

[Rubin 2002] Boris Rubin. Radon, cosine and sine transforms on real hyperbolic space. *Advances in Mathematics*, 170(2):206–223, 2002. 16, 22

[Sánchez *et al.* 2004] Gemma Sánchez, Ernest Valveny, Josep Lladós, Joan Mas, and Narcís Lozano. A platform to extract knowledge from graphic documents. application to an architectural sketch understanding scenario. In *Document Analysis Systems VI*, pages 389–400. Springer, 2004. 18

[Sarada 2014] *South African Rock Art Digital Archive*. http://www.sarada.co.za, 2014. Accessed 15 October 2014. 1

[Seidl and Breiteneder 2011] Markus Seidl and Christian Breiteneder. Detection and classification of petroglyphs in gigapixel images–preliminary results. In *VAST: International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage-Short and Project Papers*, pages 45–48. The Eurographics Association, 2011. 22, 23

[Seidl and Breiteneder 2012] Markus Seidl and Christian Breiteneder. Automated petroglyph image segmentation with interactive classifier fusion. In *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing*, page 66. ACM, 2012. 22, 23

[Seidl *et al.* 2014] Markus Seidl, Ewald Wieser, Matthias Zeppelzauer, Axel Pinz, and Christian Breiteneder. Graph-based similarity of petroglyphs. In *VISART "Where Computer Vision Meets Art"*. ECCV 2014, Springer, 2014. 21

[Seo 2008] Naotoshi Seo. *ImageClipper*, 2008. Software available at https://code.google.com/p/imageclipper/. 43

[Smith and Turner 1975] Gerald Arthur Smith and Wilson G Turner. *Indian rock art of southern California with selected petroglyph catalog*. San Bernardino County Museum Association, 1975. 18

[Sobel 1990] Irvin Sobel. An isotropic 3× 3 image gradient operator. *Machine Vision for three-demensional Sciences*, 1990. 24

[Sotak and Boyer 1989] GE Sotak and Kim L Boyer. The laplacian-of-gaussian kernel: a formal analysis and design procedure for fast, accurate convolution and full-frame output. *Computer Vision, Graphics, and Image Processing*, 48(2):147–189, 1989. 24

[Takaki *et al.* 2006] Ryuji Takaki, Junichiro Toriwaki, Shinji Mizuno, Ritsuko Izuhara, Muhiddin Khudjanazarov, and Marina Reutova. Shape analysis of petroglyphs in central Asia. *Forma*, 21(3):243–258, 2006. 17, 18

[Thomas and Wong 2004] Jim Thomas and Pak Chung Wong. Visual analytics. *IEEE Computer Graphics and Applications*, 24(5):0020–21, 2004. 23

[Van Rijssen 1987] W.J.J. Van Rijssen. Paintings in peril. *The South African Archaeological Bulletin*, pages 5–9, 1987. 12

[Viola and Jones 2001] Paul Viola and Michael Jones. Robust real-time object detection. *International Journal of Computer Vision*, 4, 2001. 2, 21, 25, 27, 28, 29, 30, 31, 32, 33, 46

[Vlachos *et al.* 2005] Michail Vlachos, Zografoula Vagena, Philip S Yu, and Vassilis Athitsos. Rotation invariant indexing of shapes and line drawings. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 131–138. ACM, 2005. 18

[Von Ahn *et al.* 2003] Luis Von Ahn, Manuel Blum, Nicholas J Hopper, and John Langford. Captcha: Using hard ai problems for security. In *Advances in CryptologyEUROCRYPT 2003*, pages 294–311. Springer, 2003. 23

[Wei *et al.* 2007] L. Wei, E. Keogh, X. Xi, and S-H Lee. Supporting anthropological research with efficient rotation invariant shape similarity measurement. *Journal of the Royal Society Interface*, 4(13):207–222, 2007. 17, 19

[Zhang and Lu 2002a] Dengsheng Zhang and Guojun Lu. Enhanced generic fourier descriptors for object-based image retrieval. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, volume 4, pages IV–3668. IEEE, 2002. 20

[Zhang and Lu 2002b] Dengsheng Zhang and Guojun Lu. Shape-based image retrieval using generic fourier descriptor. *Signal Processing: Image Communication*, 17(10):825–848, 2002. 20

[Zhang *et al.* 2013] Jiawan Zhang, Kai Kang, Dajian Liu, Ye Yuan, and Yanli E. Vis4heritage: Visual analytics approach on grotto wall painting degradations. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):1982–1991, 2013. 23

[Zhu and Keogh 2010] Qiang Zhu and Eamonn Keogh. Using CAPTCHAs to index cultural artifacts. In *Advances in Intelligent Data Analysis IX*, pages 245–257. Springer, 2010. 23

[Zhu *et al.* 2008] Qiang Zhu, Xiaoyue Wang, Eamonn Keogh, Taryn Rampley, and Sang-Hee Lee. Towards indexing and data mining all the world's rock art. *Antiquity*, 82:318, 2008. 18

[Zhu *et al.* 2009] Qiang Zhu, Xiaoyue Wang, Eamonn Keogh, and Sang-Hee Lee. Augmenting the generalized hough transform to enable the mining of petroglyphs. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1057–1066. ACM, 2009. 18

[Zhu *et al.* 2011] Qiang Zhu, Xiaoyue Wang, Eamonn Keogh, and Sang-Hee Lee. An efficient and effective similarity measure to enable data mining of petroglyphs. *Data Mining and Knowledge Discovery*, 23(1):91–127, 2011. 18, 20, 21, 23

Automatic Indexing of South African Rock Art

Amrit Purshotam
School of Computer Science
University of the Witwatersrand
Johannesurg, South Africa
amrit.purshotam@students.wits.ac.za

Turgay Celik
School of Computer Science
University of the Witwatersrand
Johannesurg, South Africa
turgay.celik@.wits.ac.za

David Pearce
Rock Art Research Institute
School of Geography, Archaeology, and Environmental Science
University of the Witwatersrand
davidp@rockart.wits.ac.za

Azizo da Fonseca
Rock Art Research Institute
School of Geography, Archaeology, and Environmental Science
University of the Witwatersrand
azizo@rockart.wits.ac.za

Hima Vadapalli
School of Computer Science
University of the Witwatersrand
Johannesurg, South Africa
hima.vadapalli@.wits.ac.za

ABSTRACT

Rock art is the archaeological term used to describe prehistoric engravings and paintings placed on natural stone and is one of the earliest known traces of modern human behaviour and cognition. Preserving these pieces of history, in the form of digital photographs, has been the mission of the South African Rock Art Digital Archive project. They have now amassed over 275 000 photographs containing a variety of geometric figures and objects such as, but not limited to, eland, elephants, rhebuck, and humans. However, only a fraction of these photographs have been classified by these objects. Thus, we present the use of the Viola Jones object detection framework combined with a support vector machine to create an automated object classifier. To test the viability of this combination, we limit our detection to just eland. The initial Viola Jones classifier was evaluated using five-fold cross-validation and was found to have a reasonable detection rate but a high false positive rate. The secondary support vector machine classifier, trained with the inclusion of these false positives, solved this problem with some success.

Categories and Subject Descriptors

I.4.8 [**Computing Methodologies**]: Image Processing and Computer Vision—*Object Recognition*

General Terms

Experimentation, Measurement, Performance

Keywords

Viola Jones, Haar, support vector machine, rock art

1. INTRODUCTION

Rock art is the archaeological term used to describe prehistoric engravings and paintings placed on natural stone and is one of the earliest known traces of modern human behaviour and cognition. Rock art is found all over the world, although South Africa is home to the oldest, with engraved ochre pieces dated to as far back as 100 000 years [12].

Rock art is a valuable source of information into the past. It allows researchers to develop an understanding of the societies, culture, and religion of those people. This is especially relevant in Africa as rock art is being used to produce a more complete history of the continent based on African perceptions rather than just colonial ones [18].

The rock art is incredibly fragile and very difficult to preserve due to its exposure to natural forces such as wind, rain, and solar radiation which is steadily and irreversibly destroying the paintings. The Rock Art Research Institute (RARI) recognised the urgent need to preserve these paintings in another form for further research and so manually digitised their copies of paintings in southern Africa. This culminated in the South African Rock Art Digital Archive (SARADA), the world's largest rock art collection consisting of over 275 000 images containing a variety of geometric figures and objects such as, but not limited to, eland, elephants, rhebuck, and humans.

The problem is that only a fraction of these images have been classified by these objects making search queries thoroughly incomplete and thus leaving most of the archive inaccessible to this type of search. In this work, we present the use of the Viola Jones object detection framework [23] combined with a secondary Support Vector Machine (SVM) [1, 4] to create an automated classifier to label the SARADA. The Viola Jones framework makes use of Haar-like features to encode knowledge about an object and, in constant time, decide whether to accept or reject a region of an image as the object. The SVM makes use of a separating hyperplane to maximise the margin between classes of data (in our case, an image with or without the object). To test the viability of this framework we limit the detection to paintings of eland, a large antelope found in southern Africa and a common figure in the rock art [19]. Examples of these paintings can be seen in Figure 2.

The rest of this paper is organised as follows. Section 2 reviews the Viola Jones object detection framework (2.1) and SVMs (2.2) as well as discuss previous work done in the area (2.3). The various problems with rock art images are described in Section 2.4. Section 3 describes our experiments as well as the results obtained. Lastly some concluding thoughts are discussed in Section 4 and

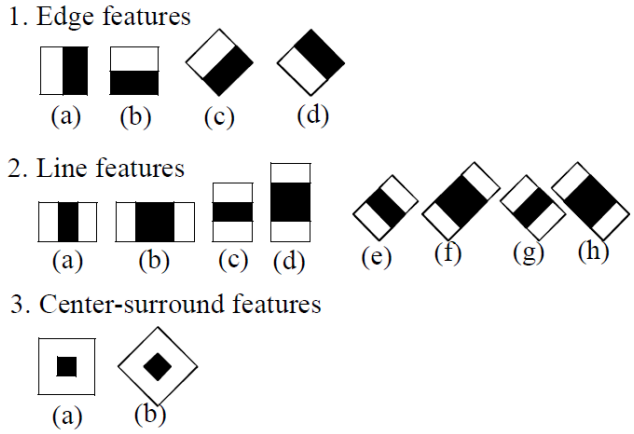


Figure 1: The extended set of Haar-like features [15]. The sum of the pixels in the white regions are subtracted from the sum of the pixels in the black regions.

details of future work.

2. BACKGROUND AND RELATED WORK

2.1 Viola Jones Object Detection Framework

Viola and Jones [23] use a boosted collection of features to classify image windows. A set of weak binary classifiers is learned from a training set using AdaBoost [11]. Each classifier is a simple function made up of rectangular sums followed by a threshold. These rectangular sums are known as Haar-like features and are simply the difference between adjacent rectangular regions in an image. The extended set [15] of features can be seen in Figure 1.

In each round of boosting, the feature with the lowest weighted error is selected. In subsequent rounds incorrectly labelled examples are given a higher weight while correctly labelled examples are given a lower weight [20].

Viola and Jones [23] noticed that within any single image, the majority of the sub-windows are negative. Hence, classification was divided into a cascade of successively more complex classifiers that each utilize more features. The idea is that the classifiers at the beginning, which use the fewest features and are thus faster, would reject most of the negative regions. The subsequent classifiers which are slower and more complex are reserved for harder regions.

2.2 Support Vector Machines

The SVM [1, 4] is a tool used to classify data. The SVM uses training data labelled with their class as input to produce a classification model which is subsequently used to classify unlabelled test data. More formally, given a training set of instance-label pairs $(\mathbf{x}_i, y_i), i = 1, \ldots, l$ where $\mathbf{x}_i \in R^n$ and $y \in \{1, -1\}^l$, the SVM requires the solution of the following optimisation problem:

$$\begin{aligned} \min_{\mathbf{x},b,\xi} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{l}\xi_i \\ \text{subject to} \quad & y_i(\mathbf{w}^T\phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0 \end{aligned}$$

Here the training vectors $\mathbf{x}_i$ are mapped on to a higher dimensional space by the function ϕ. The SVM then finds a linear separating hyperplane with the maximal margin through this higher dimensional space. $C > 0$ is the penalty parameter of the error term. Lastly, $K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)$ is the kernel function [13].

2.3 Previous Work

Little attention has been paid to object recognition in rock art by the image processing and machine learning community. To the best of our knowledge, the following paragraph summarises the body of work in this field.

The earliest such work studied the use of neural networks to classify rock art in Argentina [10]. Later Takaki et al. [22] proposed methods to characterise the shape of petroglyphs in central Asia and the properties of the group they belong to. Next, Wei et al. [24] found limited success using a time series, however its performance suffered when presented with weather-degraded petroglyphs. Zhu et al. [25] then proposed a distance measure based on the Generalised Hough Transform (GHT) to cluster southern Californian petroglyphs and also developed a tool to crowdsource the outlining of objects by humans in difficult images to aid their algorithms. Later, an automated segmentation by pixel classification of petroglyphs from Capo di Ponte, Valcamonica was presented by Seidl and Breiteneder [21]. Deufemia et al. [7] then combined Self Organising Maps (SOMs) with a Fuzzy Visual Language Parser for the recognition of petroglyphs from Mount Bego, France. Later, Deufemia and Paolino [5] leveraged the use of the Image Deformation Model (IDM) combined with SOMs. The same authors then made use of the extended Generic Fourier Descriptor for the segmentation and recognition of petroglyphs that is rotation and scale invariant [6]. Attempts have also been made for the interpretation of rock art [8, 9]

In our work, we make use of the Viola Jones object detection framework [23] combined with SVMs which is applied to a dataset of southern African rock paintings instead of just engravings or petroglyphs, the focus of the work described above. In the past the Viola Jones framework has primarily been applied to face detection [14, 15, 23] while in the field of rock art object detection only anecdotal results have been reported [16]. Support vector machines have been used in a variety of applications including, but not limited to, handwritten digit recognition, text categorisation, and object detection [2].

2.4 Background on Rock Art

Detecting objects in rock art pose several problems. The objects may vary in rotation, scale, and appearance with their features occurring in different proportions in each painting. The paintings have also deteriorated over time with some fading partially but otherwise remain intact while others have parts that have flaked off completely. Since the paintings are composed of complex scenes, varying object classes are in close proximity to each other and even overlap. Lastly, the background and rocky surface have a rough texture, which include cracks across paintings, and vary in colour.

3. EXPERIMENTAL EVALUATION

3.1 Setup

Our classification procedure is made up of a two stage cascade of the primary Viola Jones detection framework followed by a secondary SVM. The Viola Jones classifier is able to detect multiple potentially positive windows at

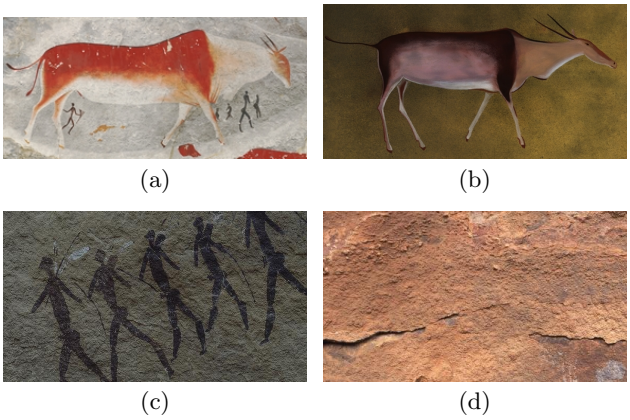


Figure 2: Examples of eland in the positive training set are shown in (a) and (b) while (c) and (d) are examples of images in the negative training set [19]

arbitrary positions and scales in an image. The SVM then has the simpler task of accepting or rejecting each of those windows since each window will have exactly one object or none at all.

To test this procedure, we limit our detection to paintings of just eland. Five-fold cross validation was chosen as the experimentation method. The Viola Jones classifiers were trained using positive training sets, 1400 to 1500 images in size, of individual eland at a side profile facing to the right and oriented at approximately 0° to the horizontal. The negative training set consisted of 1751 images and contained paintings of stick-like human figures, other animals, or just bare cave walls. Examples of each set can be seen in Figure 2. Our test sets each contained ±114 positive images and another 114 negative images. The positive images each contained one or more side profile eland facing in either direction with a maximum deviation of ±15° since it was found that this is the maximal rotational tolerance of the Viola Jones algorithm [14]. The testing images all fairly reflect the rock art domain including all the problems (except arbitrary rotations) mentioned in Section 2.4. The Viola Jones algorithm was implemented using OpenCV, an open source computer vision package [17].

Next, five-fold cross-validation was used to train the SVMs. The training sets included some of the additional false positive regions detected by the Viola Jones classifier to help minimise the false positive rate. Testing was carried out with the same testing sets as before but the eland were cropped out, flipped horizontally if needed, and scaled accordingly while the negative images included the rest of the false positive regions (i.e. not used in training) detected earlier by the Viola Jones classifier. The SVM was implemented using LIBSVM [3].

3.2 Results

Each of the five Viola Jones classifiers trained consisted of 18 stages in the cascade. As shown in Table 1, there were on average 1.50 false positive detections per image. By inspection of these false positive regions, it was concluded that they were caused by rough and jagged rocks which cast small shadows that were mistaken for features and edges. Similar looking animals were also sometimes mistaken for eland. Also seen in Table 1, the true positive rate was 67.32%. One of the difficulties in recognising eland figures was the variability in their appearance as well as their deterioration. Some of the successful detections on difficult images can be seen in Figure 3.

Table 1: Results of the five Viola Jones classifiers. FP / Img is the average number of false positives per image.

Set	# Eland	# TP	TP %	FP / Img
1	180	119	66.11	1.06
2	211	144	68.25	1.54
3	237	144	60.76	1.61
4	224	161	71.88	1.45
5	204	142	69.61	1.82
Avg	**211.2**	**142**	**67.32**	**1.50**

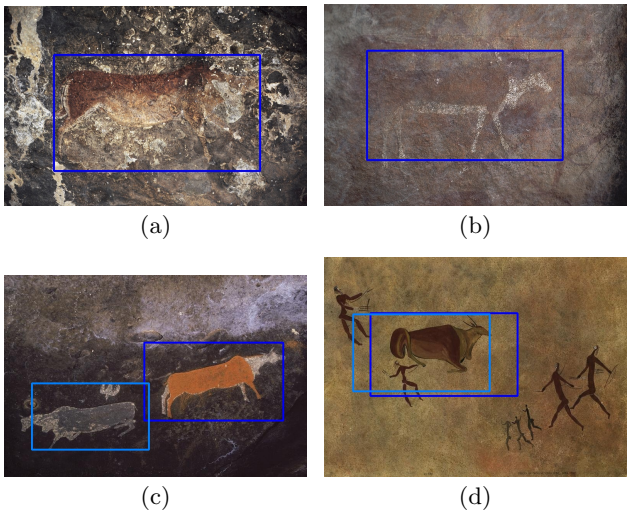


Figure 3: Examples of successful detections by the Viola Jones classifiers. They depict eland that are (a) on rough variably coloured surfaces, (b) faded, and (c) different scales and direction. (d) illustrates multiple detections on the same eland which we considered to be one detection [19].

As shown in Table 2, the SVMs helped improve the Viola Jones classifier, correctly classifying on average 73.25% and 71.27% of the positive and negative images respectively. This, in part, can be attributed to the false positive images from the Viola Jones classifier being added to the training set but also to the fact that the SVM is only required to identify one eland in an image of fixed size.

Since SVMs are used as a secondary classifier and are only able to use the potential regions of interest detected by the Viola Jones classifier as input, the overall accuracy can be estimated to have a true positive or detection rate of 49.31% (67.32% × 73.25%) at an average of 0.43 (1.50 × (100 - 71.27)%) false detections per image.

4. CONCLUSION AND FUTURE WORK

In this work we considered the use of the Viola Jones framework combined with SVMs to identify objects within the domain of rock art with the aim of using it to automatically index large archives of rock art images. Considering the extreme difficulty in detecting objects within the rock art domain due to their variability in appearance between images and their deterioration, the results obtained were

Table 2: Results of the five SVM classifiers. Since there were more true positives, the overall percentages are slightly skewed toward them.

Set	# Eland	TP %	TN %	Overall %
1	168	69.64	69.33	69.55
2	192	72.40	73.68	72.39
3	219	78.08	69.33	75.85
4	212	70.75	72	71.08
5	195	75.38	72	74.44
Avg	**197.2**	**73.25**	**71.27**	**72.66**

reasonable.

Future work will investigate alternate detection systems that use more fine grained features. Achieving rotational invariance and partial shape matching for extremely deteriorated paintings will also be explored.

5. ACKNOWLEDGMENTS

The authors would like to give special thanks to the National Research Foundation for funding (Grant 83692).

Figures 2(a), 2(b), 3(a), 3(b), and 3(c) reproduced with permission from the Rock Art Research Institute, University of the Witwatersrand, Johannesburg, South Africa.

Figure 2(c) reproduced with permission from the Hone Archive, KwaZulu-Natal, South Africa.

Figures 2(d) and 3(d) reproduced with permission from the Natal Museum, KwaZulu-Natal, South Africa.

6. REFERENCES

[1] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pages 144–152. ACM, 1992.

[2] C. J. Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.

[3] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/∼cjlin/libsvm.

[4] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[5] V. Deufemia and L. Paolino. Combining unsupervised clustering with a non-linear deformation model for efficient petroglyph recognition. In *Advances in Visual Computing*, pages 128–137. Springer, 2013.

[6] V. Deufemia and L. Paolino. Segmentation and recognition of petroglyphs using generic fourier descriptors. In *Image and Signal Processing*, pages 487–494. Springer, 2014.

[7] V. Deufemia, L. Paolino, and H. d. Lumley. Petroglyph recognition using self-organizing maps and fuzzy visual language parsing. In *Tools with Artificial Intelligence (ICTAI), 2012 IEEE 24th International Conference on*, volume 1, pages 852–859. IEEE, 2012.

[8] V. Deufemia, L. Paolino, G. Tortora, A. Traverso, V. Mascardi, M. Ancona, M. Martelli, N. Bianchi, and H. De Lumley. Investigative analysis across documents and drawings: visual analytics for archaeologists. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 539–546. ACM, 2012.

[9] V. Deufemia, V. I. Pisano, L. Paolino, and P. de Roberto. A visual analytics system for supporting rock art knowledge discovery. In *Computational Science and Its Applications–ICCSA 2014*, pages 466–480. Springer, 2014.

[10] D. Díaz and D. Castro. Pattern recognition applied to rock art. *BAR International Series*, 1016:463–468, 2002.

[11] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Computational learning theory*, pages 23–37. Springer, 1995.

[12] C. S. Henshilwood, F. d'Errico, and I. Watts. Engraved ochres from the middle stone age levels at Blombos Cave, South Africa. *Journal of Human Evolution*, 57(1):27–47, 2009.

[13] C.-W. Hsu, C.-C. Chang, C.-J. Lin, et al. A practical guide to support vector classification, 2003.

[14] M. Jones and P. Viola. Fast multi-view face detection. *Mitsubishi Electric Research Lab TR-20003-96*, 3:14, 2003.

[15] R. Lienhart, E. Kuranov, and V. Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In *In DAGM 25th Pattern Recognition Symposium*, pages 297–304, 2003.

[16] V. Mascardi, V. Deufemia, D. Malafronte, A. Ricciarelli, N. Bianchi, and H. de Lumley. Rock art interpretation within Indiana MAS. In *Agent and Multi-Agent Systems. Technologies and Applications*, pages 271–281. Springer, 2012.

[17] OpenCV. http://opencv.org/. Accessed 20 August 2014.

[18] Rock Art Research Institute. http://www.wits.ac.za/rockart/. Accessed 20 August 2014.

[19] South African Rock Art Digital Archive. http://www.sarada.co.za. Accessed 20 August 2014.

[20] R. E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. *Machine learning*, 37(3):297–336, 1999.

[21] M. Seidl and C. Breiteneder. Automated petroglyph image segmentation with interactive classifier fusion. In *Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing*, page 66. ACM, 2012.

[22] R. Takaki, J. Toriwaki, S. Mizuno, R. Izuhara, M. Khudjanazarov, and M. Reutova. Shape analysis of petroglyphs in central Asia. *Forma*, 21(3):243–258, 2006.

[23] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 4, 2001.

[24] L. Wei, E. Keogh, X. Xi, and S.-H. Lee. Supporting anthropological research with efficient rotation invariant shape similarity measurement. *Journal of the Royal Society Interface*, 4(13):207–222, 2007.

[25] Q. Zhu, X. Wang, E. Keogh, and S.-H. Lee. An efficient and effective similarity measure to enable data mining of petroglyphs. *Data Mining and Knowledge Discovery*, 23(1):91–127, 2011.