Kyffin Williams: Digital Analysis of Paintings

Final Report for CS39440 Major Project

Author: Alexander David Brown (adb9@aber.ac.uk)
Supervisor: Hannah Dee (hmd1@aber.ac.uk)

22nd April 2012
Version: 0.0.124 (Draft)

This report was submitted as partial fulfilment of a MEng degree in Software Engineering (G601)

Department of Computer Science
Aberystwyth University
Aberystwyth
Ceredigion
SY23 3DB
Wales, UK

Declaration of originality

In signing below, I confirm that:

- This submission is my own work, except where clearly indicated.
- I understand that there are severe penalties for plagiarism and other unfair practice, which can lead to loss of marks or even the withholding of a degree.
- I have read the sections on unfair practice in the Students' Examinations Handbook and the relevant sections of the current Student Handbook of the Department of Computer Science.
- I understand and agree to abide by the University's regulations governing these issues.

Signature .....

Date .....

Consent to share this work

In signing below, I hereby agree to this dissertation being made available to other students and academic staff of the Aberystwyth Computer Science Department.

Signature .....

Date .....

Acknowledgements

A large thanks to Hannah for being an amazing supervisor.

Abstract

Sir John "Kyffin" Williams was a landscape painter from Wales whose work was predominantly based in Wales and Patagonia. He studied at the Slade, one of Britain's top art schools, after epilepsy ended his career in the army during the second world war. This epilepsy made Kyffin Williams sensitive to light and is the reason his work gets darker over time [9].

Gareth Lloyd Roderick, a PhD student in the National Library of Wales, has collected data such as the date or location, of these paintings. This data allows for some interesting analysis; particularly that of temporal or geographical classification of a given painting. That is being able to take a painting and decide the year or location in which it was painted from a database of existing, known, works by Kyffin Williams.

Temporal analysis will be the focus of this project as it allows for a diverse range of techniques; from statistical analysis of colour values of the paintings to looking at the length and style of paintbrush strokes. Geographical analysis would likely be very difficult, especially as the locations depicted were often sketched on-site then painted in a studio.

Whilst it would be nice to be able to predict the age of a painting with no known year, it is far more interesting to try to guess the year of paintings for which the date is known. This project will use leave-one-out cross-validation to help measure the effectiveness and validity of the analysis techniques employed in this project. Leave-one-out validation can be used with this project as the data set is small enough not to incur large performance overheads and the overall speed of the program is irrelevant so long as it completes within a decent amount of time.

One major limitation with this is it also includes the technique used for classification, some techniques might work better with K-Nearest Neighbour whilst other might benefit from more complex methods of classification. This means I will either have to stick with a single classification algorithm and hope it's a good one for all techniques. Or I could find the best machine learning technique for each individual analysis technique, then perform comparison. This does assume that the best machine learning technique for every analysis technique exists within the scope of the project.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Background & Objectives

1.1 Sir John "Kyffin" Williams

Sir John "Kyffin" Williams (1918-2006) was a Welsh painter and printmaker, widely regarded as the defining artist of Wales during the 20th century [7]. He was advised to take up art by a doctor after failing a British Army medical examination because of an 'abnormality' (epilepsy) as something which would not tax his brain.

He studied at the Slade School of Fine Art and taught art in Highgate School, after which he retired to Anglesey until he died in 2006 after a long battle with cancer.

His most characteristic pictures are of Welsh landscapes, painted with thick layers of oil paint applied with a palette knife [5]. Most of his paintings are highly textural; to the point of being 3-dimensional.

As his life progressed Kyffin's 'abnormality' grew steadily worse, especially when exposed to bright light. As a result most of his paintings are of overcast Welsh landscapes and tend to become visibly darker over time [9]. By eye it is generally quite easy to approximate the time period in which a painting was created.

In 1969 he won a scholarship to study and paint in Y Wladfa; the Welsh settlement in Patagonia. This period of his life is very obvious from his paintings as there is a complete contrast in colour between Patagonian and Welsh landscapes.

1.2 Interdisciplinary work with the National Library of Wales

This project was initially suggested through a conversation between Hannah Dee and Gareth "Llyod" Roderick about image processing and art. Llyod is a PhD student at the National Library of Wales (NLW) researching (TODO: Find out what Llyod's thesis title is). Their initial idea was to be able to geolocate a Kyffin painting on a map to build up a geographical representation of Kyffin's work.

Hannah started to create a prototype for performing geographical analysis, this proved to be a difficult task and one which is still being researched.

However, the nature of Kyffin's illness and painting style allows for a second form of analysis: temporal. As previously stated it is fairly easy to judge by eye a good approximation of the period in which a Kyffin painting was created. It should, therefore, follow that this process can be performed digitally.

1.2.1 Future Work

1.3 Existing Work

1.3.1 Edge-Orientated Gradients

1.3.2 Brush-stroke Analysis

1.4 Analysis Objectives

1.4.1 Colour-space Analysis

1.4.2 Texture Analysis

1.4.3 Brush-stroke Analysis

1.5 Classification Objectives

1.5.1 Classification

1.5.1.1 Use of Weka

1.5.1.2 Learning Classifier Systems (LCS)

1.5.2 Exemplars

Chapter 2

Development Process

2.1 Introduction

2.2 Modifications

Chapter 3

Design

3.1 Overall Architecture

3.2 Some detailed design

3.2.1 Even more detail

3.3 User Interface

3.4 Other relevant sections

Chapter 4

Implementation

4.1 Python

4.1.1 Python setuptools

4.2 Image Processing

4.2.1 OpenCV

4.2.2 scipy and numpy

4.2.3 Edge Orientation

4.3 Classification

4.3.1 K-Nearest Neighbour

4.3.2 Weka 3

4.3.2.1 Attribute-Relation File Format (ARFF)

Chapter 5

Testing

5.1 Overall Approach to Testing

5.2 Validation

5.2.1 Leave-One-Out Cross Validation

5.2.2 Validation using Weka

Chapter 6

Evaluation

6.1 Evaluation of Requirements

6.2 Evaluation of Design

6.3 Evaluation of Tools

6.3.1 Python

6.3.1.1 setuptools

6.3.2 OpenCV

6.3.3 Weka 3

6.3.3.1 Attribute-Relation File Format (ARFF)

6.3.4 scipy and numpy

Appendices

0.1 Equations

0.1.1 Statistical Equations

0.1.1.1 Mean

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

0.1.1.2 Standard Deviation

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{2}$$

0.1.2 Distance Equations

0.1.2.1 Manhattan Distance

$$d_1(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|_1 = \sum_{i=0}^{n} |p_i - q_i| \tag{3}$$

0.1.2.2 Euclidean Distance

$$d_1(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=0}^{n}(q_i - p_i)^2} \tag{4}$$

0.1.3 Filter Equations

0.1.3.1 Gradient Direction

$$\theta = \operatorname{atan2}\left(\frac{\delta f}{\delta x}, \frac{\delta f}{\delta y}\right) \tag{5}$$

0.1.3.2 Gabor Filter

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda} + \psi\right) \tag{6}$$

where:

$$x' = x\cos\theta + y\sin\theta$$
$$y' = x\sin\theta + y\cos\theta$$

Annotated Bibliography

[1] P. Azad, "The Integrating Vision Toolkit (IVT)," Website, 2011, used documentation (http://ivt.sourceforge.net/doxygen/) and examples (http://ivt.sourceforge.net/examples.html) to create a simple image blurring application to test the capabilities and ease of use of the library. Used these resources from the 16 October 2012 to the 24 October 2012. [Online]. Available: http://ivt.sourceforge.net/

[2] J. Bacardit and X. Llorà, "Large-scale data mining using genetics-based machine learning," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3, no. 1, pp. 37–61, Jan. 2013, a paper recommended by Julie Greensmith for information on a Learning Classifier System (LCS) which Julie believes will yield good results with the Kyffin Williams project. [Online]. Available: http://dx.doi.org/10.1002/widm.1078

[3] I. E. Berezhnoy, E. O. Postma, and H. J. van den Herik, "Authentic: Computerized Brushstroke Analysis," in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE, July 2005, pp. 1586–1588, defines a method of analysing brushstrokes by applying a circular filter across a digital image to pick up the ridges of a brushstroke. This can then be used to pick out individual brushstrokes in order to be able to fit a nth order polynomial to them. Though this paper focuses on authenticating Van Gogh's paintings, it could easily be applied to the work of Kyffin Williams and may allow for some interesting analysis. [Online]. Available: http://dx.doi.org/10.1109/icme.2005.1521739

[4] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000, used Python (http://opencv.willowgarage.com/documentation/python) and C++ (http://opencv.willowgarage.com/documentation) documentation for library reference and some learning on image processing/computer vision. Used since 11 October 2012.

[5] I. Chilvers, J. Glaves-Smith, and I. Chilvers, *A dictionary of modern and contemporary art*. Oxford University Press, 2009. [Online]. Available: http://www.worldcat.org/isbn/0199239665 0199239665.

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, ser. CVPR '05, vol. 1. Washington, DC, USA: IEEE, June 2005, pp. 886–893 vol. 1, describes a method of producing histograms of edge orientations which may prove to be a useful analysis technique for Kyffin Williams' art. The most interesting part of this paper is the use of segmentation and binning of gradients which seems like it could be useful to differentiate different parts of the image which may be painted in different styles. [Online]. Available: http://dx.doi.org/10.1109/cvpr.2005.177

[7] J. Davies and A. Gymreig, *The Welsh Academy encyclopaedia of Wales*. University of Wales Press, 2008, pp. 957–958. [Online]. Available: http://www.worldcat.org/isbn/9780708319536 9780708319536.

[8] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, Nov. 2009, citation for the Weka data mining software. Weka is a Java based tool which can be used to run a lot of classifiers to a dataset, making it a very useful tool to apply to the Kyffin Williams project. Weka allows the application of complex machine learning techniques without having to spend a lot of time learning, understand and implementing said techniques. [Online]. Available: http://dx.doi.org/10.1145/1656274.1656278

[9] R. Harris, "How Rolf learnt to paint like Sir Kyffin Williams," BBC Broadcast, Feb. 2011, a video on the BBC by Rolf Harris about some of Kyffin Williams' life and about his interesting style of painting. [Online]. Available: http://www.bbc.co.uk/programmes/p00f6nyt

[10] J. Li, L. Yao, E. Hendriks, and J. Z. Wang, "Rhythmic Brushstrokes Distinguish van Gogh from His Contemporaries: Findings via Automated Brushstroke Extraction," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 6, pp. 1159–1176, June 2012, defines a complex method for analysing individual brush strokes which has been used to classify the period of two paintings by Van Gogh. This technique could be very powerful when applied to Kyffin Williams' work. This could be one of the most important techniques for the whole of the Kyffin Williams project. [Online]. Available: http://dx.doi.org/10.1109/tpami.2011.203

[11] S. J. D. Prince, *Computer vision : models, learning, and inference*. Cambridge University Press, 2012, learning reference for Computer Vision and Machine Learning. [Online]. Available: http://www.computervisionmodels.com/ 9781107011793.

[12] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Y. Tinevez, D. J. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, "Fiji: an open-source platform for biological-image analysis." *Nature methods*, vol. 9, no. 7, pp. 676–682, July 2012, used as part of referencing FIJI documentation to test out of capabilities and ease of use of the library. [Online]. Available: http://dx.doi.org/10.1038/nmeth.2019

[13] SciJava, "Aggregator project for the Fiji plugins 2.0.0-SNAPSHOT API," used as reference documentation to create a simple application to blur an image. Used from 11 October 2012 to 24 October 2012. [Online]. Available: http://fiji.sc/javadoc/

[14] D. Stork and M. Duarte, "Computer Vision, Image Analysis, and Master Art: Part 3," *IEEE Multimedia*, vol. 14, no. 1, pp. 14–18, 2007, defines a potential method of analysing regions in a painting, this could be interesting to apply to Kyffin Williams' work as the regions in his earlier work are a lot less well defined as those in his later work. [Online]. Available: http://dx.doi.org/10.1109/mmul.2007.6

[15] D. G. Stork, "Computer Vision and Computer Graphics Analysis of Paintings and Drawings: An Introduction to the Literature Computer Analysis of Images and Patterns," ser. Lecture Notes in Computer Science, X. Jiang and N. Petkov, Eds. Berlin, Heidelberg: Springer Berlin / Heidelberg, 2009, vol. 5702, ch. 2, pp. 9–24, notes a lot of useful literature to look

at and some useful terminology and ideas for analysis techniques too. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-03767-2_2