Can we date an artist's work from catalogue photographs?

Alexander David Brown
Department of Computer Science,
Llandinam Building,
Aberystwyth University,
Aberystwyth,
Ceredigion,
SY23 3DB
Email: adb9@aber.ac.uk

Gareth Lloyd Roderick
Email: glr7@aber.ac.uk

Hannah M. Dee
Department of Computer Science,
Llandinam Building,
Aberystwyth University,
Aberystwyth,
Ceredigion,
SY23 3DB
Email: hmd1@aber.ac.uk

Lorna M. Hughes
Email: lorna.hughes@llgc.org.uk

***Abstract*—Kyffin Williams, art changes over time, blah blah blah. Features, colour, edges, histograms of oriented gradients; strong correlation using leave-one-out methodology. Exemplars; artistic and statistic.**

I. Introduction

This paper presents a interdisciplinary computational study into the modelling of artistic style, and how this style changes over time. The artist Sir John (Kyffin) Williams painted from X to 2004, and produced many landscape paintings during this period – he was a prolific painter. His style evolved from a very figurative, representational style, to something more abstract: the computer scientists on our team would say that the paintings became more *blocky*; the art historians that *XXXwhatever lorna wants to say*. Through a collection of digital photographs of oil paintings, collected from museum websites, catalogues and other sources, we first investigate whether it is possible to date a painting from an unknown year based upon image features alone.

II. Background

III. The image dataset

Our image dataset consists of 325 paintings, with associated metadata. Metadata includes title, year or year ranges, genre, original painting size, painting materials and image size.

These pictures are challenging: they are not colour calibrated; some suffer from reflections (towards the end of his life Kyffin painted using exceptionally thick and textural strokes, which gives specularities on the catalogue images); they are at varying resolutions; and come from a range of different cameras. Image size bears little relation to the original painting size, and some images are even optimised for the web. Table I below summarises the dataset

Type	Number	Number (Known date)	Notes
Landscapes	247	64	
Portraits	52	35	
Seascapes	11	2	
Still lifes	4	1	
Other	8	0	

TABLE I. A summary of the Kyffin Williams painting dataset used

IV. Methodology

Within our database of X paintings, we know the actual year of painting for Y artworks. In order to determine the accuracy of our results we have used a leave-one-out cross validation methodology. This involves us taking a painting for which we know the year, and then guessing that year; thus we are able to tell whether we are right, and if we are wrong, exactly how wrong we are.

To simplify the classification stage we use a K-Nearest Neighbour classifier with the other Y-1 paintings for which we know the date; we expect that there may be some underlying pattern in the change of style but for this work have concentrated on features for classification rather than classification or regression itself.

Thus for each feature set, we take all paintings for which we know the year of creation; select one painting, and find its nearest neighbours within that feature space. We also know that painting's actual year, and we can plot actual against predicted year for all known-year paintings. To measure goodness of fit, the Pearson's product-moment correlation coefficient was calculated on these orderings; this provides us with a performance measure of each classifier.

V. An exploration of colour and texture features

The digital analysis of paintings is quite a wide area to look into; there are many feature spaces which can be used; from colour-space, edge detection, texture analysis and brush-stroke recognition.

Initially we focused on simple colour-space analysis; taking the mean RGB across each painting and running it through a simple classification algorithm, k-nearest neighbour. Other colour models, such as HSV, were also used to compare and contrast against RGB.

Colour histograms of the paintings were also produced as a more accurate measurement of colour space. We continued to use k-nearest neighbour, but with different methods of calculating distance between each point in feature space; namely chi squared.

As a lot of Kyffin Williams' paintings are highly textural, edge detection and texture analysis were thought to be good techniques to explore.

Edge detection involved applying one of the various edge detection algorithms available, applying it to each painting. The distance measure is base on the number and strength of edges in the painting.

Texture analysis is a continuation of edge detection. Instead of just taking the strength and number of edges; it creates a histogram of orientated gradients[1] begin to build up a more faithful representation of the texture of a painting. To begin with steerable filters at 0, $\frac{\pi}{4}$, $\frac{\pi}{2}$ and $\frac{3\pi}{4}$ were applied to the image to bin by those orientations.

$$g_e(x) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x}+\frac{y^2}{\sigma_y}\right)} \cos(2\pi\omega_{x_0}x + 2\pi\omega_{y_0}y) \quad (1)$$

Where

$(\omega_{x_0}, \omega_{y_0})$ Defines the centre frequency

and

(σ_x, σ_y) the spread of the Gaussian window

Gabor filters were also used with a greater range of angles to produce a more accurate representation of the texture of the painting.

Another method for producing histograms of orientated gradients is to apply two discrete derivative masks to the image to get the gradient of x and y and then to work out the gradient direction at each point. This can then have a histogram created from it to provide a different representation of the texture of the image.

VI. Exemplars: Can We Improve Results by Incorporating Expert Knowledge?

For exemplar-based classifiers the result of the classification was a new point in space, the statistically-chosen exemplar, which was then compared to an existing point in space, the artistically-chosen exemplar. The performance measure applied to each classifier was the squared error of the distance from the statistically-chosen exemplar to the artistically-chosen exemplar.

The data on artistically-chosen exemplars opened up the options for different methods of classification. Rather than using a k-nearest neighbour algorithm to classify each point in the feature space we could just take the year of the nearest exemplar to that point and use that for classification.

This then raised the question as to what the digitally-chosen exemplars would be. Statistically exemplars would be the centroid of a group of paintings in feature space; a relatively simple operation to perform digitally.

A. *Comparing Artistic and Statistical Exemplars*

The same distance measure used to run k-nearest neighbour on the feature-based classifiers was used to generate the error between the statistical and artistic exemplars.

VII. Conclusion

Acknowledgment

The authors would like to thank...

References

[1] 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, Jun. 2005, pp. 886–893 vol. 1. [Online]. Available: http://dx.doi.org/10.1109/cvpr.2005.177