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. Sir John Kyffin Williams (1918-2006) was one of the predominant figures in Welsh art of the twentieth century. Kyffin- as he was almost universally known in Wales- studied at the Slade School of Art and worked as an art master at Highgate School, before returning to live on his native Anglesey in 1973. He was a prolific painter and once claimed to have painted 'two pictures per week when in London, and three per week when in Wales.'[1, p.209] With a career spanning from the mid-1940s to approximately 2004, this rate amounts to a large body of work.

His style evolved from a very figurative, representational style, to something more expressive, which retained representational qualities: the computer scientists on our team would say that the paintings became more *blocky*; the art historians that his landscapes are almost constructed with swathes of textural paint. His was a style characterised by thick impasto paint, applied almost exclusively with palette knife, although the application technique appears to change over time.

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. We show that using a K-nearest neighbour classifier, tested using a leave-one-out methodology, we can obtain a strong correlation between image feature descriptors and year of painting. We go on to investigate whether exemplar based methods are able to improve on this, using what we call artistic exemplars (paintings selected by an expert as being typical for a particular year) and statistic exemplars (paintings which are near the centre of year-based clusters in feature space).

II. Background

Although also a portrait painter, Kyffin Williams is primarily known for his landscape paintings of north west Wales and Anglesey. While his technique and style changed over the years, his landscapes in oil are instantly recognisable, often featuring bold chunks of colour, and various points during his career bold black outlines to figures and landscapes features. Greens, browns and greys often form the palette of his paintings of the Welsh landscape. These colour selections seem appropriate for the artists claim that melancholy, derived from the 'dark hills, heavy clouds and enveloping sea mists,' is a 'national characteristic of the Welsh.[1] This combination of colour selection and technique seems appropriate for the depiction of the areas where he painted. Many of his most successful paintings are said to have a 'dark quality' in depicting 'rain lashed hillsides,' and it was this darkness which 'makes his landscapes so distinctively Welsh.'[2] Both of the above descriptions could be applicable to paintings in the collections of the National Library of Wales such as Snowdonian Summits (1970-1990) or Yr Wyddfa a Grib Goch (1950-1960).

The aesthetic of Kyffins Welsh landscapes is contrasted by the paintings he made following a trip to Patagonia to paint the landscape and people of the Welsh communities there in 1968 as part of a Winston Churchill Foundation scholarship. The colours and application of paint in pictures produced in following this journey (such as Lle Cul, Henry Roberts, Bryngwyn Patagonia, Euros Hughes Irrigating his Fields, all 1969, National Library of Wales) differ starkly from paintings of Welsh landscapes, incorporating pinks, purples and oranges. This contrast, combined with the fact that the Patagonian pictures were produced during a definite period of time has reiterated our interest in how analysing the formal qualities of pictures from different collections remotely, using digital images can be used to apply a date of production to works where the date is unknown.

Kyffin Williamss work in well represented in public collections in Wales (particularly at the National Library of Wales, the National Museums and Galleries of Wales and Oriel Ynys Mn, Anglesey). His pictures, often depicting the landscape and people of north-west Wales were also tremendously popular with the art buying public. Of the 325 paintings by Williams in public collections in the UK listed on the BBC/Public Catalogue Foundations Your Paintings website, 212 of them are in the collections of the National Library of Wales[3]. Many of these paintings were bequeathed to the Library as part of a larger bequest by the artist (including works on paper and other archival material). Many of the pictures which came to the library from the artists studio had little in the way of metadata, and as such have been catalogued with large date-ranges estimating the dates of production.

A. A digital humanities approach to art history

XXX We need some words for this bit

B. Computer vision and the analysis of paintings

XXX Alex or Hannah are going to have to read [4], cite it sensibly, and pull some other references which are relevant

When we consider computer vision-based analysis of painterly style we find that the vast majority of work concentrates on brush stroke detection and analysis. For example, Berezhnoy and colleagues in [5] detect brush-strokes by moving a circular filter across the whole painting to find the ridges of strokes, then filling any unbroken areas. They then shrunk these areas to a single pixel line and fitted a n^{th} order polynomial to this line.

Li et al [6] use a combination of edge analysis and clustering in colour space to determine strokes; a number of heuristics involving branching, stroke-width modeling, and gap filling are then used to refine the original brush stroke estimates. One interesting element of this work, from our perspective, is the ability to date some of Van Gogh's paintings to a known period in his career.

Techniques based upon stroke analysis, whilst applicable to the work of some artists, are not applicable to all. In particular, Kyffin Williams painted with a pallet knife and whilst there are clear "*strokes*" identifiable in his style, these vary widely in size and shape, so the morphological techniques which can detect strokes in Van Gogh's work are unlikely to pay off when considering the blockier Williams oeuvre. Another difference of note is that much work on computerised painting analysis (including [6], [5]) is based upon high resolution scans acquired in controlled conditions, whereas the current paper deals instead with a collection of photographs from catalogues, websites, and other disparate sources.

III. The Image Dataset

Our image dataset consists of 325 paintings, with associated metadata. Metadata includes title, year or year ranges (for those works where year is unknown but can be estimated by curators), genre, original painting size, painting materials and image size.

These photographs of paintings are challenging in and of themselves: 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	Genre unknown or studies

TABLE I. A SUMMARY OF THE KYFFIN WILLIAMS PAINTING DATASET USED

XXX It may be worth putting in something here about image size vs painting size?

IV. Methodology

Within our database of 325 paintings, we know the actual year of painting for 102 artworks. In order to determine the accuracy of our results, rather than work with the full dataset (and work with images with uncertain metadata in the form of date ranges), we have used a leave-one-out cross validation methodology. This involves us taking a painting for which we know the year, and then using our classifier to guess that year; thus we are able to tell whether we are right. We are also able, if we are wrong, to determine exactly how wrong we are.

To simplify the classification stage we use a K-Nearest Neighbour (KNN) classifier with the other 101 paintings for which we know the date. KNN is a fast, non-parametric classifier which makes no assumptions about the underlying patterns in the data, merely that paintings from around the same time will be similarly located in our feature space(s). Whilst we suspect that there may be some broader underlying trend in the change of style, for this work have concentrated on features for classification rather than the question of 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. The year assigned by our classifier to that painting is the mean of the K neighbours. Figure 1 provides an overview of this classification methodology.

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. It is also possible to test Pearson's R for statistical significance; thus significance levels are reported alongside R in this paper. With all of the feature spaces we consider, it is possible treat the painting descriptors as histograms. This allows us to use a single distance measure, namely chi-squared, in our K-nearest neighbour classification.

V. An Exploration of Colour and Texture Features

The digital analysis of paintings is a broad reseach area. Within the methodology we have selected, there are many feature spaces which could be useful: from simple analysis of the way in which colour changes over time, through edge detection, to texture analysis. Within this work we have

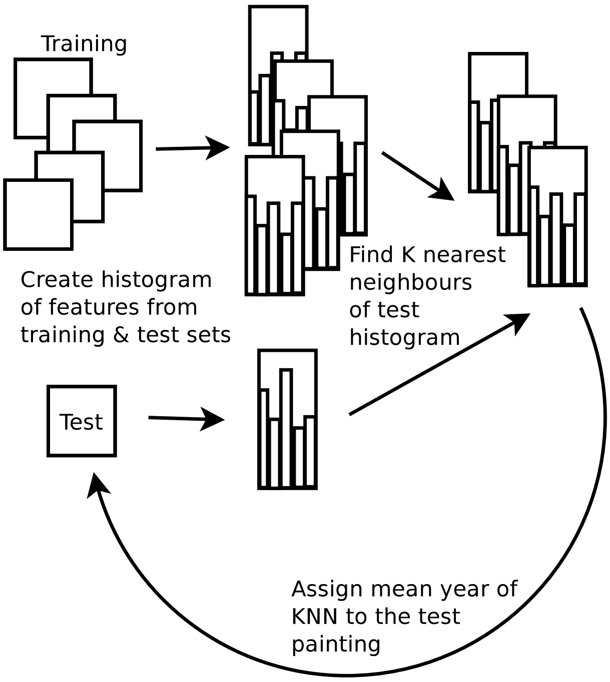

Fig. 1. Overview of the classification methodology

concentrated on lower level image features – colours, textures, and edges – rather than attempt to extract brush strokes. Kyffin Williams painted with a pallette knife rather than a brush, and his work is characterised by angularity rather than identifiable "strokes".

There is a clear (to the eye) trend in colour usage, as the paintings get "gloomier" over time, so we started with simple colour-space analysis: taking the mean RGB for each painting and using this with our KNN classfier; we also tested other colour spaces, such as HSV. Promisingly this provided us with a positive correlation.

Staying with the colour variation theme, we then used colour histograms, which provide a more precise representation of the way Kyffin Williams used colour. These histograms were developed by counting the number of pixels within a particular colour range for each painting, and then building a normalised histogram representing the colour usage.

As a lot of Kyffin Williams' paintings are highly textural, edge detection and texture analysis were thought to be good techniques to explore. To investigate the edge properties of the artworks we apply a Canny [7] edge detector to the paintings, and then build a histogram based upon counts of edge strength.

Texture analysis is a continuation of edge detection. Instead of taking simply the strength and number of edges, we create a histogram of orientated gradients as in [8]. In this way we begin to build up a richer representation of the texture of a painting. Given the change in style of Kyffin Williams' work, moving away from figurative representations with curved lines towards more blocky rectilinear brush strokes, we expect these edge orientation frequencies to change over time. To this end we used simple steerable filters S, applied to the image at 0, $\frac{\pi}{4}$, $\frac{\pi}{2}$ and $\frac{3\pi}{4}$.

$$S\left(\frac{\pi}{2}\right) = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \tag{1}$$

Equation 1 shows a sample steerable filter, in this case $S(\frac{\pi}{2})$, the filter which gives the highest response when presented with horizontal lines. By convolving each image with filters tuned to different orientations, we can build a histogram recording the frequency of lines at each orientation.

Fig. 2. Example image – Coastal Sunset, 1990-2006

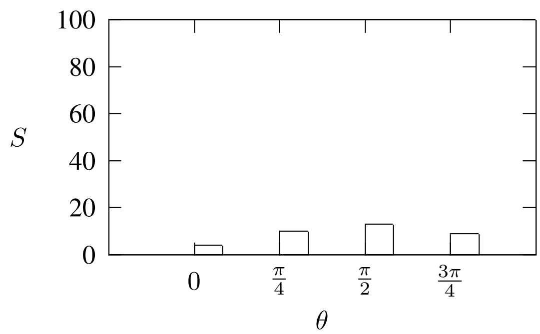

Fig. 3. Steerable filter strength $S(\theta)$ on the example image in figure 2

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

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

Where $(\omega_{x_0}, \omega_{y_0})$ defines the centre frequency, and (σ_x, σ_y) the spread of the Gaussian window.

XXX Hannah or Alex to look at a textbook for a better version of the above scary maths and cite.

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 richer representation of the texture of the image.

XXX Alex is going to combine the various correlation-against-K graphs so that we have one graph with all techniques on the same axes

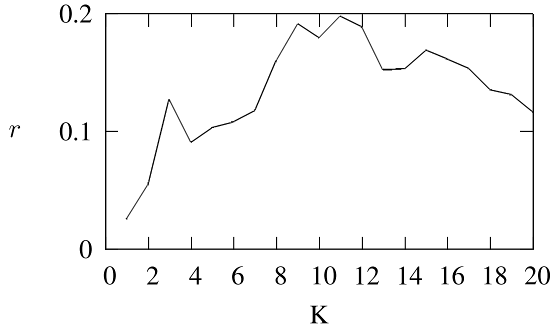

Fig. 4. Correlation Coefficients r against K values for K-Nearest Neighbour on RGB Colour-Space Analysis

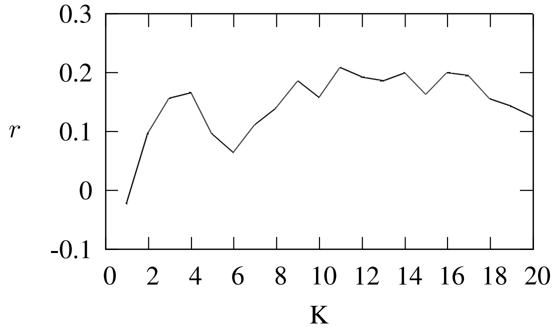

Fig. 5. Correlation Coefficients r against K values for K-Nearest Neighbour on HSV Colour-Space Analysis

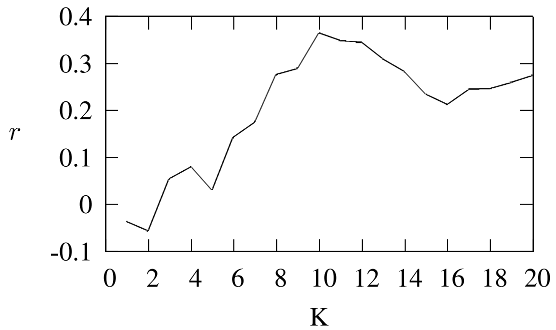

Fig. 6. Correlation Coefficients r against K values for K-Nearest Neighbour on Histogram Analysis

The one parameter of our classifier is the choice of K in K-nearest neigbour. Simply setting $K = 1$ has the effect of assigning the year of the nearest painting in feature space to the current test painting, whereas setting $K = 102$ has the effect of giving each painting the mean value of the entire dataset. Clearly a point between these two extremes would be best; from Figure 8 we can see that for many of the feature spaces we consider, the optimum K value is around 10.

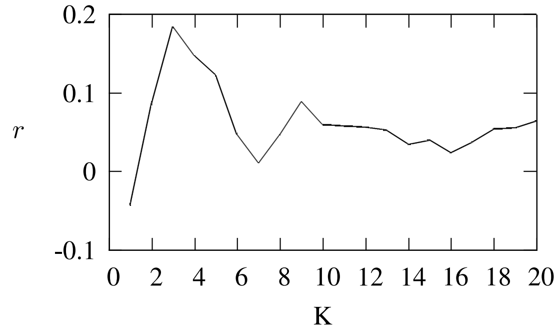

Fig. 7. Correlation Coefficients r against K values for K-Nearest Neighbour on Edge Strength Analysis

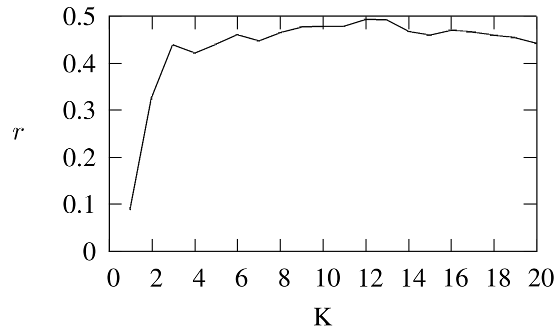

Fig. 8. Correlation Coefficients r against K values for K-Nearest Neighbour on HOG Analysis

XXX Alex - do we have HOG results for steerable, Gabor, and derivative masks?

XXX Alex - do we have any ensemble methods? IIRC combining HOG with HSV gave us impressive correlations

VI. Exemplars: Can we improve results by incorporating expert knowledge?

We have also investigated the utility of incorporating expert knowledge within our framework. For each year represented in our collection we asked Dr Paul Joyner, head of the purchasing and donations unit at the National Library of Wales, board member of the Trustees of the Kyffin Williams estate who has written widely on Welsh art and Kyffin Williams, to choose the one painting which best represents the artist's work for that year. These we consider to be connoisseurially/artistically chosen exemplars (*artistic exemplars*, for short), which we can then use as a representation of that particular year.

The data on artistic exemplars opens up the options for different methods of classification. Rather than using K-nearest neighbour to classify each point in the feature space, we take the year of the nearest exemplar to assign that year to the painting in question.

This then raises the question of whether we can determine *statistical exemplars* to compare with our *artistic exemplars*, and if so what the digitally-chosen exemplars would be. We can either use cluster centres (which provides us with a point in feature space which will not correspond to an actual painting), or the nearest actual painting to the feature space centroid for

Technique	r	$P(r)$
RGB	0.179	0.05757
HSV	0.158	0.09518
HSV Histogram	0.364	0.00007
Edge Strength	0.060	0.52877
HOG	0.479	0.00000008

TABLE II. CORRELATION COEFFICIENTS

a year. The former technique does not, strictly speaking, give us an exemplar; the latter chooses as exemplar the painting which best represents a particular year according to a particular feature space.

XXX Alex - We need results:-)

XXX Alex again - can we get an indication of any situations where our statistic exemplars match the artistic ones, and the converse situation (artistic exemplars nowhere near centres of feature space)? This is a good place for us to include a couple of extra photos of paintings

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. CONCLUSIONS AND FUTURE DIRECTIONS

To the best of our knowledge this is the first work that attempts to date work by an artist by year. Similarly, we believe we are the first to try and perform digital analysis of paintings from a range of catalogue and web images. The results presented here show that computer vision *can* help with the job of dating art within an artist's body of work.

Future directions will involve testing the methods presented here on the works of other artists who have shown great stylistic variation over the course of their career: we would like to build a dataset of, for example, David Hockney works. Whilst we have not yet performed this test we are hopeful of success; by avoiding brushstroke detection (which we expect to be artist specific) we hope to have developed techniques with application across a broader range of artistic styles.

ACKNOWLEDGMENT

The authors would like to thank...

REFERENCES

[1] K. Williams, *Across the Straits: An Autobiography*. Gomer, 1993.

[2] J. Davies and C. Cymru, *100 Welsh heroes*. Culturenet Cymru, 2004.

[3] "Your Paintings - Kyffin Williams," Mar. 2013. [Online]. Available: http://www.bbc.co.uk/arts/yourpaintings/paintings/search/painted_by/kyffin-williams_artists

[4] D. G. Stork, "Computer vision and computer graphics analysis of paintings and drawings: An introduction to the literature," 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.

[5] I. E. Berezhnoy, E. O. Postma, and Herik, "Automatic extraction of brushstroke orientation from paintings," *Machine Vision and Applications*, vol. 20, no. 1, pp. 1–9, 2009.

[6] 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, Jun. 2012.

[7] J. F. Canny, "A computational approach to edge detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[8] 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.