Evolutionary Automated Recognition and Characterization of an Individual's Artistic Style

Taras Kowaliw, Jon McCormack, and Alan Dorin

Abstract— In this paper, we introduce a new image database, consisting of examples of artists' work. Successful classification of this database suggests the capacity to automatically recognize an artist's aesthetic style. We utilize the notion of Transform-based Evolvable Features as a means of evolving features on the space, these features are then evaluated through a standard classifier. We obtain recognition rates for our six artistic styles — relative to images by the other artists and images randomly downloaded from a search engine — of a mean true positive rate of 0.946 and a mean false positive rate of 0.017. Distance metrics designed to indicate the similarity between an arbitrary greyscale image and one of the artistic styles are created from the evolved features. These metrics are capable of ranking control images so that artist-drawn instances appear at the front of the list. We provide evidence that other images ranked as similar by the metric correspond to naïve human notions of similarity as well, suggesting the distance metric could serve as a content-based aesthetic recommender.

I. Introduction

In this paper we use evolutionary computation to discover a useful collection of features for the description of individual artists' styles. We first introduce a novel and public database of images: the samples are panels from practising artists in the genre of comics and graphic novels, and thus are a non-representational sample of each artist's work. Our work's success is evaluated by two means: correctly classifying images by originating artist; and contrasting the performance of a generated distance metric to our naïve expectations of the properties of similar styles.

Our approach can be summarized as follows: Firstly, we define a collection of features well-suited to distinguishing artists' styles using an evolutionary feature creator (EFC). This technique uses genetic programming (GP) to create pixel-level functions which define transforms on the space of grayscale images, transforms ideally useful for characterizing important sub-patterns of an image; Next, we define a feature space and distance metric based on this discovered collection of features; Finally, we explore that distance metric's behaviour on a control set of images.

There are two obvious applications of this technology. Such a classifier or distance metric could be used to troll the web in search of examples of an artist's work. This might serve as a means of detecting copyright infringement; The use of low-level pixel-based features here suggests that even partial sections of work might be recognized if pieces have been incorporated into larger wholes.

The authors are with the Faculty of Information Technology, Monash University, Clayton VIC 3800, Australia (web: http://www.csse.monash.edu.au/~cema, email: taras@kowaliw.ca)

Secondly, and more interestingly, a successful distance metric would return not only instances of an artist's work, but also instances of work by artists with a similar style. Such a ranking metric might appeal to users more interested in purely aesthetic characteristics, rather than the more semantic results returned by association or keyword matching. There is need for this sort of measure: For instance, Adomavicius and Tuzhilin note that in the design of recommender engines — systems which suggest database entities to users based on some evidence of likely utility — current content-based techniques are limited by the typically sparse features available [?]. Or, consider the case of content-based image retrieval, where images are suggested based on how closely they correspond to previously selected images, as measured by (some approximation of) the human visual system. Datta *et al.* point out that here too appropriate features are often unknown, and that automated feature extraction methods are difficult to apply to multimedia data. Further, that in some cases visual similarity might be more critical than semantic similarity [?].

Finally, this work shows that evolutionary feature creation can be extended successfully to another field of image recognition, demonstrating the breadth of the approach.

II. Review

Since image spaces tend to involve very large numbers of dimensions, there is need for intelligent means of dimensionality reduction. Feature extraction, the mapping of images to numeric description vectors, is an important field in computer vision. Linear techniques can be shown to approach optimality by a measure of inter- and intra-class spread, but sometimes make inefficient assumptions regarding the distribution of data and the form of solution; As a result, there is much work in the nonlinear generation of features [?], [?].

The complexity of the tasks involved make evolutionary computation a natural means of approaching the problem. Early evolutionary techniques utilized LISP-style programs to control a roving agent [?], [?]. More recent techniques tend towards the use of a collection of low-level descriptions of image features evaluated via a GP-derived mathematical expression [?], [?], or as a composition of image processing filters [?]. To varying levels, the assumption that the data sources can be well represented by a small number of low-level features, rather than by the larger raw pixel-space in which images are defined, makes use of GP in image classification similar to the more general feature construction

found in a broader spectrum of machine learning tasks [**?**], [**?**].

Transform-based Evolvable Features (TEFs) were first introduced in [**?**], and further developed in [**?**]. TEFs operate on the raw-pixel space, and use GP to define transforms on the space of greyscale images. These transforms, ideally, serve to highlight important characteristics of the database at hand. The explicit purpose of a TEF-based evolutionary feature extraction system is to automatically define features for a given image database when the relevant patterns are complex or unknown. To date, TEF-based approaches have achieved state-of-the-art classification accuracies in an object recognition database and a database of medical cell images indicating a form of muscular dystrophy [**?**].

There have been several attempts to use automated techniques to describe or classify artistic styles (we do not here consider artist authentication, since this work tends to have differing goals and utilizes non-image sources, such as surface texture and x-ray scans). All have relied on the use of pre-defined image features, using some classifier to predict either individual artist or broader artistic genre or both, where images have been mined from databases of "great" artists. The most complete comparisons are by Zujovik *et al.* [**?**] and by Shen [**?**], who both considered collections of existing and novel features evaluating a collection of paintings by famous artists. Shen achieves a recognition accuracy of approximately 0.69. Zujovik *et al.* achieved recognition rates of approximately 0.68 on binary genre classification; Further, their recreation of previous features showed, in some cases, far inferior results to the original reported success rates, suggesting great variance due to data source. Standardized sources are obviously of merit here.

We see several potential issues with the above approaches which our current work sidesteps:

- the great artists have been selected by art theorists or historians for their own purposes. This introduces several potential biases which may have consequences for classifiability, not the least of which is the possibility that the "greats" have been selected precisely due to their uniqueness, or their divergence from previous art practice. The artists in the database we employ (CAPD, See Section **??**) were chosen in a randomized (but non-representative[1]) fashion; Each is a contemporary commercial artist working in a single popular genre. As a result, this database seems more likely to mimic the difficulties in the classification of an individual artist's style as it would occur in many practical scenarios.
- we introduce a control group, one based on results from a search engine, which can be considered an unbiased source in terms of the classifiability of the samples;
- we do not rely on pre-defined features, allowing the particularities of sample distinction to emerge from the artistic practice, rather than being framed in previous expectation.

[1]Indeed, retrospectively we see that the majority of artists are Canadian, an obvious bias from the collector.

III. A DATABASE OF ARTISTS' WORK

Our Comic and Graphic Novel Artist Panel Database (CAPD; freely available at http://kowaliw.ca/capd) is a collection of grayscale panel drawings organized by artist, and a control group. The images do not depict any consistent object, character, nor anything else representational; Instead, they are examples of an artist's aesthetic style (for the duration of the project included in the source material, at least). All are examples of the comic and graphic novels genre, implying grayscale images containing highly stylized caricatures.

The CAPD consists of seven classes of image, hereafter labelled $C = \{A\&B, BNS, Cats, Flower, Love, Other, Words\}$, where *Other* is the control group. Examples of each class are shown in Figure **??**. The control group was generated by randomly selecting a subset of images returned by a search engine when queried for black-and-white line drawings corresponding to the keywords "comic graphic novel". There are an additional twenty constructed images in the control group, designed to represent extreme points in image space (e.g. all white, simple gradients, noise, etc.). There are 150 training images and at least 80 validation images for each class, with 240 validation images in the control group. Each image is 8-bit greyscale of size 200×200 pixels or less.

We define the $fitness$ of some given classifier, σ, on a set of images from our database, S, to be

$$fitness(\sigma, S) = \prod_{c \in C} (1 - FPR(\sigma, S, c)) \qquad (1)$$

where FPR is the achieved false positive rate. When computing $fitness$ for training sets of images, we utilize a 50/50 train-test split; For evaluating validation sets of images, we use 10-fold cross-validation. For each run we utilized 40 randomly selected images per class for training and 80 images per class for final validation. Note that $fitness$ is stochastic due to the randomized ordering and folding of presentation of images. We should also point out that $fitness$ is a more difficult measure than is required for application: firstly, it demands the creation of a multi-classifier, whereas a binary classifier should be sufficient for any particular usage; secondly, it demands correct classification of the *Other* class, which is likely unnecessary. Regardless, optimization of $fitness$ should lead to good values of many other interesting success measures.

IV. TRANSFORM-BASED EVOLVABLE FEATURES

Here we provide an overview of TEF-based feature extraction; The interested reader may consult [**?**] for details. Our overall system is an evolutionary feature creator (EFC), a system which evolves individuals which define a collection of features for a provided image database. The goal of an individual in the EFC is to transform the images to a numeric vector suitable for classification.

An individual is composed of a collection of TEFs, or individual transforms on the space of grayscale images.
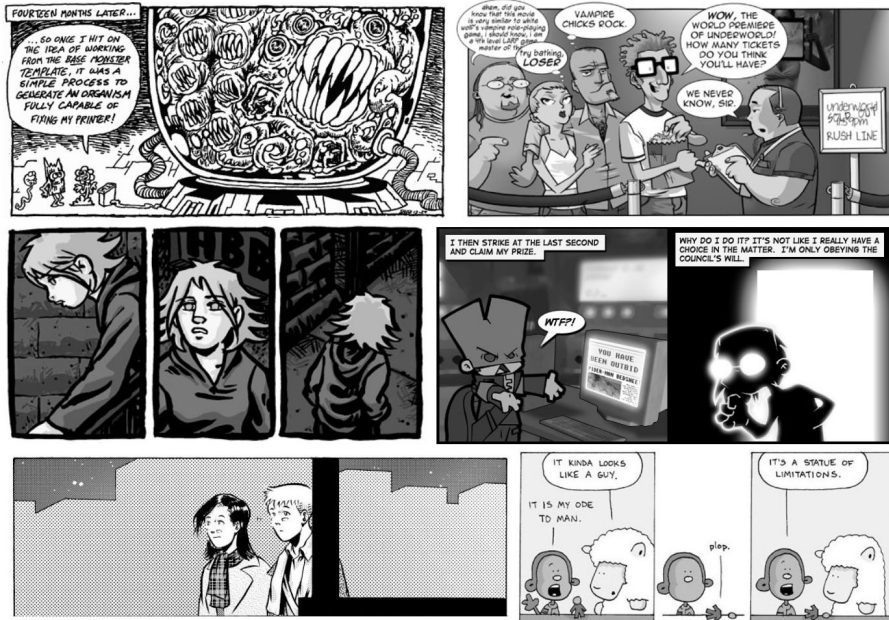
Fig. 1. Examples of artists' work. From left to right, top to bottom: S. Notley's **Flower** (http://www.angryflower.com); R. Perez & R. Coughler's **BNS** (http://www.butternutsquash.net); B. Rivers' **Words** (http://benjaminrivers.com/emptywords); S. Ramsoomair's **Cats** (http://www.vgcats.com); E. Kim's **Love** (http://www.inkskratch.com); J. Burgess' **A&B** (http://www.jimburgessdesign.com). All drawn images, here and throughout the paper, used with permission and © the originating artists. B. Rivers' work is protected under the creative commons.
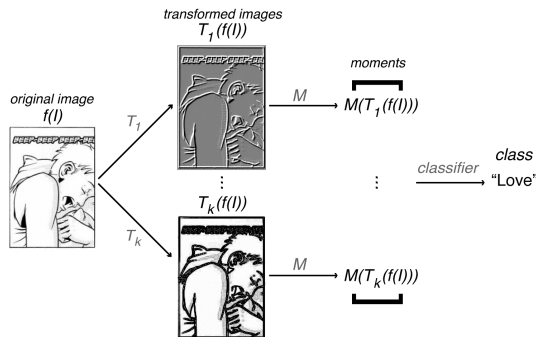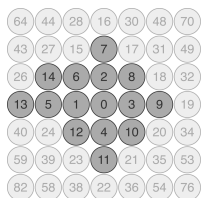

Fig. 2. Overview of the TEF-based approach to evolutionary feature extraction.

3. The *VN* neighbourhood is one of several means of parameterizing two-dimensional discrete space about a central point. Highlighted here is *VN-15*.

These TEFs are GP graphs defined on the space of pixel neighbourhoods, which, when applied as a sliding window over the image, return a transformed image. The transformed image is then converted to a set of numeric values via some moment functions, and utilized in the individual's resulting feature vector. In this case, we use a single moment, the normalized $0, 0$-th geometric moment, $M_{00} = \frac{1}{|I|} \sum_{p \in I} f(p)$.

Each individual is evaluated by training a 1-NN classifier and computing the individual's fitness. The use of a classifier during fitness evaluation sometimes leads to features which do not generalize to other classifiers; However, use of the classifier also effectively eliminates assumptions regarding the distribution of data points following transform, allowing us to use a wide space of potential transforms.

An overview is shown in Figure **??**.

V. EXPERIMENTS IN THE RECOGNITION OF ARTISTS' STYLES

As in previous work, an evolutionary algorithm was used to find individuals maximizing fitness. After an informal parameter search consisting of approximately 25 runs with randomly selected parameter values, we settled on a parameter set believed to be ideal for the problem at hand. We use a population size of 30 (doubled in the initial generation); selection via a tournament of size 3; a graph size of 250; a maximum number of transforms of size 25, and maximum initialization number of transforms of 5; a variable-size neighbourhood of type *VN* with a maximum size of 36 (see Figure **??**); and probabilities of crossover, mutation, merger, and pruning of 0.33, 0.02, 0.02, and 0.2. The pruning probability was purposefully set higher than that dictated by mere selection for accuracy in an attempt to generate parsimonious solutions. Ten runs were conducted with the above parameters, achieving a mean validation $fitness$ of 0.843 (s.d. 0.04).

A. The Critic, and the Definition of Feature Space

The best individual from the best run was extracted, and will hereafter be termed *the critic*, since it is through this individual's features that future images will be evaluated. The critic attained a validation fitness of 0.868. The accuracies for each class, in the binary problem consisting of distinguishing between a particular artist from the remaining artists and

the *Other* class, are shown in Table **??**, attaining a mean true positive rate of 0.946 and a mean false positive rate of 0.017 for each class, when evaluated as a binary classification problem. It should be pointed out that these features were chosen to optimize a multi-classifier, and it is possible that features evolved specifically to maximize the recognition of some artist's work in a binary classifier might achieve better results. Generally speaking, the 1-NN classifier does best at maximizing the true positive rate, unsurprising since this was the classifier used during the evolution of the features. Other classifiers, however, also do reasonably well on this feature set.

The critic utilizes a total of sixteen transforms, defined on a neighbourhood of type *VN-15*. The value of these transforms as features, as determined by Information Gain Ranking (on the multi-classifier), ranges from 0.97 to 0.21, where three features are excluded as completely non-discerning. These transforms — excluding three non-discerning transforms, with intra-transform redundancy removed, and in order of information gain — are:

$$T_0 = \min \left\{ thresh \left(i_3^{i_2}, i_2 \min\{i_0, i_{10}\} \right), \right.$$
$$\left. \min\{i_6, i_{13}\} + thresh(\max\{i_2, i_{14}\}, \max\{i_2, i_{10}\}) \right\}$$

$$T_1 = thresh \left(i_7 + (i_{10} - i_{12}), \sqrt{\left(\frac{i_9}{i_{14}} \right)^{i_0}} \right)$$

$$T_2 = \frac{i_4 - \left(i_7 + \frac{i_1}{i_8} \right)}{-193.76}$$

$$T_3 = \frac{(i_7 + i_{11})^{i_5}}{244.82}$$

$$T_4 = \sqrt{\frac{i_{11}}{i_3}}$$

$$T_5 = \min\{i_9, i_{14}\}$$

$$T_6 = (i_7)^{i_2} \cdot (i_{10})^{\max\{i_0, i_9, i_{11}\}}$$

$$T_7 = i_0 i_6 - \min\{i_{10}, i_{13}\}$$

$$T_8 = thresh \left((i_9)^{i_8}, \frac{i_{10}}{(-144.43)^{i_6}} \right)$$

$$T_9 = \min \left\{ i_1 + thresh(\min\{i_3, i_{14}\}, i_{10}), \frac{145.04}{i_{14}} \right\}$$

$$T_{10} = i_5 i_7 - (i_2 + i_{14})$$

$$T_{11} = thresh \left(\frac{\min\{\min\{i_{14}, (i_{13} - i_8)\}, i_9\}}{i_4}, i_8 \right)$$
$$\cdot thresh \left(thresh(i_1, i_9), (i_{13})^{i_7} \right)$$

$$T_{12} = thresh \left(thresh(i_8, i_6), thresh(i_3, \min\{i_0, i_1\}) \right)$$

The action of some of these transforms on sample images is shown in Table **??**. We will write these 13 features as $\Phi(f(I)) = (M_{00}(T_0((f(I))), ..., M_{00}(T_{12}((f(I))))$. Hence, $\Phi(f(I))$ refers to the location of an arbitrary grayscale image in the feature space defined by the critic, hereafter called simply the feature space. Below we outline some of these features and the characteristics they highlight.

The transform T_1 causes a sharp change in output when the central pixel is close to zero, and hence can distinguish be-
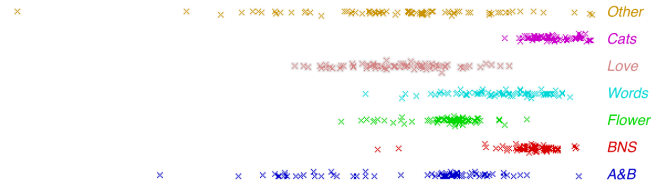

Fig. 4. Spread of values by the T_1 feature.

tween pure white and near white. Light greys are highlighted, while pure whites are left dark (note the difference between action on images in the *Love* as opposed to the *Flower* class). The classes *BNS* and *Cats*, using the most detailed backgrounds, are sent almost entirely to white. Again, *Love* is distinguished due to the use of fine cross-hatching, which is acted upon differently than both flat grey and less regular photographic backgrounds (see Figure **??**). This feature also has a tendency to highlight artifacts of JPEG compression.

The transform T_2 is a very sensitive detector for horizontal changes. It is activated when $\frac{i_1}{i_8} \approx 193$, meaning that a pixel on the right of the neighbourhood must be very close to black, and a pixel on the left must be fairly bright. This feature is fairly adept at distinguishing both *Flowers* and *Words* from most other images, as these two classes tend to contain thick black lines cutting across flat light backgrounds.

Transforms T_2 and T_4 are similar, both using division to detect changes in the horizontal and vertical directions, respectively. Transform T_2 is more specific, using division by a constant to select for a narrow band of values. This narrow range excludes, for instance, black text on a white background. This narrow range is especially adept at distinguishing the *Words* category, which has a large number of dark-grey to black transitions.

Like transform T_2, transform T_3 also uses a constant to select a narrow range of values. Pixel values are converted to one of only three possible output values, loosely corresponding to light, mid, and dark ranges, save that unvaried black is mapped to a lighter colour. Hence, T_2 seems capable of measuring the overall darkness of an image discounting unvaried fields of pure black, possibly able to identify the parts of an image that indicate the mood of the panel (dark greys) as opposed to structural elements (written text, boxes, strokes).

This same strategy seems to be recovered by transform T_6, which also maps fields of pure black to white, but then also blurs the remaining image (through the maximum of the central, right-most and bottom-most pixel intensities), a further measure to discount sharp, dark edges.

The transform T_{10} detects bright values along the diagonal, dominated by a preference for pure darkness along the vertical axis. We believe this combination rarely occurs with line drawing, and is instead more common with natural textures (e.g. from photographic sources). Hence, T_{10} serves as a means of detecting the photographic elements found in the *Cats* class. Like transform T_1, this transform also highlights pure white as opposed to near-white; Hence, comparison of this feature alongside T_1 will yield the overall darkness

THE BINARY CLASSIFICATION SUCCESS RATES ON THE VALIDATION IMAGES (560 INSTANCES) BY CLASS.

classifier	A&B		BNS		Cats		Flower		Love		Words		**mean**	**best**
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
1-NN	**0.950**	0.029	**0.925**	0.033	**0.975**	0.008	0.913	0.023	0.850	0.023	**0.988**	**0.010**		
Naïve Bayes	0.950	**0.009**	0.863	0.060	0.963	0.071	**0.925**	0.046	**0.913**	0.210	0.925	0.033	0.946	0.017
Decision Tree	0.813	0.031	0.825	**0.025**	0.925	**0.006**	0.825	0.027	0.700	0.042	0.938	0.013		
Log. Regression	0.863	0.029	0.850	**0.025**	0.925	0.017	0.888	**0.017**	0.713	**0.033**	0.963	**0.010**		

TABLE II

ACTION OF A SELECTION OF THE CRITIC'S TRANSFORMS ON SAMPLE IMAGES FROM THE CAPD. TRANSFORMS T_2 AND T_3 HAVE BEEN SUBJECTED TO CONTRAST STRETCHING TO MAKE FEATURE VALUES MORE VISIBLE.
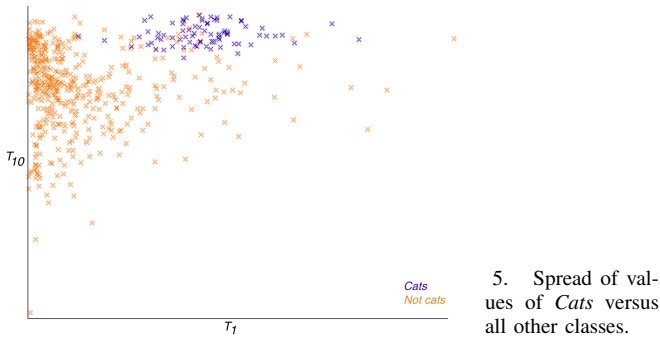
Cats
Not cats

5. Spread of values of *Cats* versus all other classes.

of the image, spread by the source (photographic, flat, or cross-hatched) of that darkness. The spread of values, in conjunction with transform T_1, is shown in Figure **??**.

There are several features which behave like edge-detectors, especially transform T_4. Unsurprisingly, these edge-detectors tend to favour the horizontal axis, as to be expected with images containing western text. Most of the edge-detectors are highly specialized and incorporated into larger features, and hence do not behave like typical edge-detectors, such as the Sobel or Laplacian filters. Other transforms detect cross-hatching and text features. The existence of a JPEG artifact-detecting transform is worrisome, since it suggests that non-aesthetic data source information might be influencing the recognition accuracy. In the next section, we will provide evidence that this concern can be overlooked.

VI. ESTIMATING THE DISTANCE BETWEEN AN ARBITRARY IMAGE AND AN ARTIST'S STYLE

In this section we attempt to create a distance metric between an arbitrary image and an artist's style. We will do so by: (a) mapping the provided image into the previously discovered feature space; and (b) computing an estimate of the probability that the feature values obtained indicate membership in the class associated with that artist from the CAPD data instances.

We originally attempted to use a robust, non-parametric distance metric (specifically Mahalanobis distance) to define the distance between points in space. However, based on our informal visual estimation, simple Euclidean distance was a better measure of human-intuitive similarity; Indeed, the classifier used to evaluate the evolving features used Euclidean distance as a similarity metric [**?**], meaning that an appropriate weighting of features is likely implicitly inserted in the feature definitions themselves by the evolutionary algorithm. Hence, we define the distance between two images, x and y as $d(x,y) = \sqrt{\sum_{0 \leq i \leq 12}(M_{00}(T_i(x_i)) - M_{00}(T_i(y_i)))^2}$.

Next, we attempt to estimate the probability that a given image is in some particular class. Given the relatively large number of dimensions in our space, it is likely impossible to demarcate any reasonably sized region and compute the frequency of samples within that region, without obtaining zero nearly everywhere: this is the familiar "curse of dimensionality" that plagues machine learning. Instead, we will use the distance between the current point and the closest

training points in the CAPD (since we know from recognition experiments that distance measures can be meaningful on this space). Since we wish to avoid undue influence of outliers, we will use several nearest neighbours in the calculation.

Let us choose a particular class c from the classes represented in the CAPD, where $\{c_i\}$ are the associated training instances. Then, given some new point p, we will choose three points from c minimizing $d(c_i, p)$, and refer to these points as *3-NN*(p, c). Finally, we will define the "class-ness" or c-ness of p to be

$$c\text{-ness}(p) = \max\{d(c_i, p) | c_i \in \textit{3-NN}(p, c)\} \qquad (2)$$

The c-ness measure will be large any time there are not at least three instances from the training set nearby; It will be small when there are many instances nearby (ideally excluding influence by outliers). Hence, we expect that c-ness is strongly inversely correlated with the desired probability measure, and hence, that we can use c-ness as an approximation.

A. Experiments with c-ness

240 search engine-derived *Other* validation images were used for experimentation with the *c-ness* measure. For each class, a collection of 20 images from the associated artist's validation set were selected randomly and were added to the *Other* images. These images were ranked, and examined for two properties: Firstly, two measures corresponding to the capacity of the *c-ness* measure to choose samples of an artist's work from a random collection of images. Secondly, we considered the top unrelated images returned by the ranking for similarity to the artist's style, a notion alas far less easily measured.

Firstly, we computed the relative *c-ness* for each of the 260 validation images. The number of true artist-drawn images returned within the top 20, and the median ranks are:

class	no. in top 20	median rank (of 260)
A&B	19	9.5
BNS	17	10.5
Cats	16	10.0
Flower	14	12.0
Love	13	10.5
Words	15	12.5
mean	**15.7**	**10.8**

These results suggest that the distance metric preserves the efficacy of the recognition process, allowing for a ranking procedure to place true artist-drawn instances at the front of the list.

We now turn to a discussion of the capacity of *c-ness* to perform as a similarity metric across several artists' work. This claim contains both a social and a subjective component, and is thus difficult to evaluate. We will, however, present some evidence that might convince the reader of the value of the measure.

Using *c-ness*, we can compute the mean distance between classes, along with the closest images. For instance, we may write that the *BNS-ness* of *Flower* is the mean distance between an image from the *Flower* class to the set *BNS*. Table **??** shows the mean distance between the CAPD classes.

From this data, we can conclude that *A&B* is most similar to *Love*, and vica versa. This corresponds to expectations, since both use light colours, thin sharp edges, and large fields of white or grey. *BNS* is most similar to *Cats*, and vica versa; also roughly in line with expectations, since both use highly detailed and photographic backgrounds, and thin, sharp edges. *Words* is somewhat similar to both *BNS* and *Cats*, likely due to the usage of darker colours and large fields of black and dark grey.

Surprisingly, the two styles which make prodigious use of thick, black lines, *Flower* and *Words*, are not particularly similar to each other. Note that the feature T_2 tends to return nearly black images, making the relative weighting of the measure small under Euclidean distance (and, consequently, 1-NN classification). Hence, the distance metric here states that feature T_2 is significant enough to measure, but less significant than other features. Indeed, *Flower* is not particularly similar — according to *c-ness* — to any of the other classes.

Within the validation images retrieved from the search engine there existed, by chance, an image by R. Coughler, one of the artists of *BNS*. While the subject matter was similar to the CAPD source materials, the image itself was a scanned sketch, and hence likely created via differing technique than the series. By *BNS-ness*, this image ranked 28th. Given that eighteen of the preceding images came from the *BNS* class, we can consider this image as ranking 10th of 240 randomly-selected images.

We also note that the majority of images ranked within the top twenty, in terms of *Love-ness*, were self-described examples of the Manga style; This was not true of any other style. We note that Kim's *Love as a Foreign Language*, a story involving a foreigner's experiences in Korea, is advertised as a Manga.

Finally, we note that although none of the CAPD artists registered as particularly close to *Flower*, there did exist an artist — J. Brown (http://jeffreybrowncomics.blogspot.com) — in the *Other* class who returned three images within the top twenty-five lowest according to *Flower-ness*. The artist's style (see Figure **??**), at least as reflected in the particular works included in the *Other* class, contains several stylistic features similar to S. Notley's: near exclusive use of black-and-white; thick strokes; an avoidance of straight lines or right angles; an expressive child-like writing style; and textures composed of small lines.

VII. Summary

Here we have described the use of an evolutionary feature extraction technique, applied to the recognition of artistic styles. GP was used to evolve a collection of pixel-based transformations on the space of comic and graphic novel panel images suitable for classification. This is further evidence that the TEF approach to feature evolution is a highly robust approach, capable of self-adapting to several different image databases.

We have shown that high accuracies are attainable, relative to images by other artists in the database and relative to images randomly selected from search engine results. Further,

Fig. 6. Examples of S. Notley's (*top*) and J. Brown's (*bottom*) images. Images © S. Notley and J. Brown resp., reprinted with permission.

we have shown that a distance metric can be constructed from these evolved features, one which allows for a ranking of a collection of images in such a way that recovers the validation accuracy of the classifier. Hence, we can describe the distance between an arbitrary grayscale image and some given artistic style.

There is evidence that our derived similarity metric can be used as a means of filtering search results to include previously unseen but human-recognizably similar artistic works. We have shown instances in which use of the *c-ness* measure: (a) groups similar artistic styles together in an intuitive fashion; (b) is capable of discovering works by the same artist from different contexts; (c) is capable of returning images from a similar sub-genre; and (d) selected a previously unseen artist with similar aesthetic qualities. From this, we suggest that a similar system could be used to create a content-based aesthetic similarity measure for use in recommender engines or content-based image retrieval systems.

Obvious future directions involve the similarity metric: (a) our approach is naïve — it was selected on the basis of informal experimentation — and there may exist more analytically motivated choices; and (b) there is recent work on similarity in visual art that might serve as a basis for meaningful comparison [**?**].

References

[1] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. on Knowl. and Data Eng.*, 17(6):734–749, 2005.

[2] D. Aha and D. Kibler. Instance-based learning algorithms. *Machine Learning*, 6:37–66, 1991.

[3] D. Andre. Automatically defined features: The simultaneous evolution of 2-dimensional feature detectors and an algorithm for using them. In K .E. Kinnear, editor, *Advances in Genetic Programming*. MIT Pres, 1994.

[4] M.A. Bustos, M.A. Duarte-Mermoud, and N.H. Beltrán. Nonlinear feature extraction using fisher criterion. *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)*, 22(6):5–32, 2009.

	A&B	BNS	Cats	Flower	Love	Words
A&B-ness	**0.062**	0.414	0.612	0.320	**0.149**	0.577
BNS-ness	0.416	**0.090**	**0.156**	0.281	0.403	0.228
Cats-ness	0.525	0.126	**0.102**	0.340	0.527	0.254
Flower-ness	0.254	0.284	0.366	**0.096**	**0.228**	0.290
Love-ness	**0.145**	0.288	0.457	0.238	**0.105**	0.423
Words-ness	0.460	**0.170**	0.203	0.263	0.393	**0.113**

[5] R. Datta, D. Joshi, J. Li, and J.Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, 40(2):1–60, 2008.

[6] D. Graham, J.D. Friedenberg, D.N. Rockmore, and D.J Field. Mapping the similarity space of paintings: image statistics and visual perception. *Visual Cognition*, in press, 2009.

[7] H. Guo, L. Jack, and A. Nandi. Feature generation using genetic programming with application to fault classification. *IEEE Trans. on Systems, Man and Cybernetics — Part B: Cybernetics*, 35(1):89–99, 2005.

[8] T. Kowaliw and W. Banzhaf. The automated generation of image features from genetic programming-based transforms. *TBD*, TBD:TBD, 2010 *(In Submission, contact corresponding author for preprint)*.

[9] T. Kowaliw, W. Banzhaf, N. Kharma, and S. Harding. Evolving novel image features using genetic programming-based image transforms. In A. Tyrrell, editor, *CEC'09: Proceedings of the Eleventh conference on Congress on Evolutionary Computation*, pages 2502–2507, 2009.

[10] J. Koza. Simultaneous discovery of detectors and a way of using the detectors via genetic programming. In *IEEE International Conference on Neural Networks*, volume 3, pages 1794–1801, 1993.

[11] K. Krawiec. Genetic programming-based construction of features for machine learning and knowledge discovery tasks. *Genetic Programming and Evolvable Machines*, 3(4):329–343, 2002.

[12] Y. Lin and B. Bhanu. Evolutionary feature synthesis for object recognition. *IEEE Trans. on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, 35(2):1094–6977, 2005.

[13] J. Shen. Stochastic modelling western paintings for effective classification. *Pattern Recognition*, 42:293–301, 2009.

[14] S. Shirakawa, S. Nakayama, and T. Nagao. Genetic image network for image classification. In *Applications of Evolutionary Computing*, volume 5484 of *LNCS*, pages 395–404. Springer-Verlag, 2009.

[15] A. Song and V. Ciesielski. Texture segmentation by genetic programming. *Evolutionary Computation*, 16(4):461–481, 2008.

[16] Feng Zhang. A polygonal line algorithm based nonlinear feature extraction method. In *Data Mining, 2004. ICDM '04. Fourth IEEE International Conference on*, pages 281–288, 2004.

[17] J. Zujovic, L. Gandy, S. Friedman, B. Pardo, and T. N. Pappas. Classifying paintings by artistic genre: An analysis of features &classifiers. In *in Proc. IEEE Workshop on Multimedia Signal Processing*, 2009.