X.Zhang, G.Nagy, Style comparisons in Calligraphy, Procs. SPIE/EIT/DRR, San Francisco, Jan. 2012.

# Style comparisons in calligraphy

Xiafen Zhang[*a], George Nagy[†b]

[a]College of Information Engineering, Shanghai Maritime University, Shanghai, China 201306
[b]Electrical, Computer, and Systems Engineering, RPI, Troy, NY, USA 12180

## ABSTRACT

Calligraphic style is considered, for this research, visual attributes of images of calligraphic characters sampled randomly from a "work" created by a single artist. It is independent of page layout or textual content. An experimental design is developed to investigate to what extent the source of a single, or of a few pairs, of character images can be assigned to the either same work or to two different works. The experiments are conducted on the 13,571 segmented and labeled 600-dpi character images of the CADAL database. The classifier is not trained on the works tested, only on other works. Even when only a few samples of same-class pairs are available, the difference-vector of a few simple features extracted from each image of a pair yields over 80% classification accuracy for a same-work vs. different-work dichotomy. When many pairs of different classes are available for each pair, the accuracy, using the same features, is almost the same. These style-verification experiments are part of our larger goal of style identification and forgery detection.

**Keywords:** style verification, style identification, font recognition, forgery detection, calligraphy

## 1. INTRODUCTION

Calligraphy is an essential part of our cultural heritage. Outstanding exemplars – some over one thousand years old, others almost contemporary – are preserved in museums throughout the world and digitized versions are becoming accessible through the Web. While no language has a monopoly on beautiful writing, the largest collections are of Chinese, Indic, Persian/Arabic, and Latin scripts. Some specimen command attention primarily because of their content or historic significance, others are treasured because of their delightful combinations of shapes. The most admired works define the intersection of literature and fine arts.

This study explores style differences in a recently compiled database of Chinese calligraphic characters. The database contains 13,351 segmented and labeled character images digitized at 600 dpi by the China Academic Digital Associative Library (CADAL)[1], which manages the China-US Million Book Digital Library Project, and is an important part of the Universal Digital Library (UDL)[2]. Each character image is indexed by the Dublin Core bibliographic entry for the source book, page number, the name (and number) of the specific work, its GB 2312 code (*label*), and a unique character identifier (*CID*). A *work* here is an instance of a calligraphic composition by an artist. Nine books and about 50 works account for about two-thirds of the data. About a third of the characters are from white-on-black photographs of stone rubbings. The digitization and data collection process, and detailed statistics about the contents of the CADAL Calligraphic Database, were reported at ICDAR 2011 [3].

We attempt to determine whether two character images, or two groups of character images, are from the same work or not. We use the word *style* to describe the commonality of shapes found in an individual work that differentiates it from other works. Our styles are therefore more specific than the broad taxonomy of *Great Seal, Small Seal, Clerical, Regular, Running,* and *Cursive* styles. Some of the works in the database may be considered intermediate between two of these. A few defy classification into the conventional categories.

Determining whether a character image is from the same work as another seems easier if the two images represent the same character, i.e., if they have the same label. Often, however, It is of interest to determine commonality of source when same-label characters are unavailable.

[*] xfzhang@shmtu.edu.cn
[†] nagy@ecse.rpi.edu

The methods applicable to both problems are similar to those developed for font recognition[4], writer classification[5], biometric verification[6], and forgery detection[7]. In biometrics, verifying a subject's claimed identify from a fixed spoken or written phrase, is called *strong enrollment* or *strong training*. Verification that does not require the same phrase at enrollment time as at verification time and must therefore rely only on voice or script similarity rather than direct comparison, is dubbed *weak training*[8]. Initial experiments on forgery detection in Chinese calligraphy were presented by one of the current authors[9]. Calligraphic style models were described by Zhang and Zhuang[10]. Our work differs from these in the choice of features, classifiers, and in the exploitation of a larger calligraphic database that allowed development of a sound experimental design based on randomized sampling.

The chosen approach is based on pair dichotomy. The objects to be classified consist of *pairs* of patterns. There are two classes: Class 1, s*ame-work pairs*, and Class 2, *different-works pairs*. We adopt a statistical classification paradigm, with a randomly selected training set consisting of an equal number of same-work pairs and different-works pairs, and a test set consisting also of these two kinds of pairs. In same-label pair classification, the training and test patterns are *same-work-same-label* (SWSL) and *different-works-same-label* (DWSL) pairs. In different-labels pair classification, they are *same-work-different-labels* (SWDL) and *different-works-different-labels* (DWDL) pairs. Same-label experiments cannot demonstrate style-awareness, because they would work just as well if the work commingled different styles for each label. Only different-labels experiments can reveal the difference between *style* features and *character-shape* features. The classifier is trained on differential style-sensitive features. Fig. 1 shows examples of the four kinds of character image pairs.

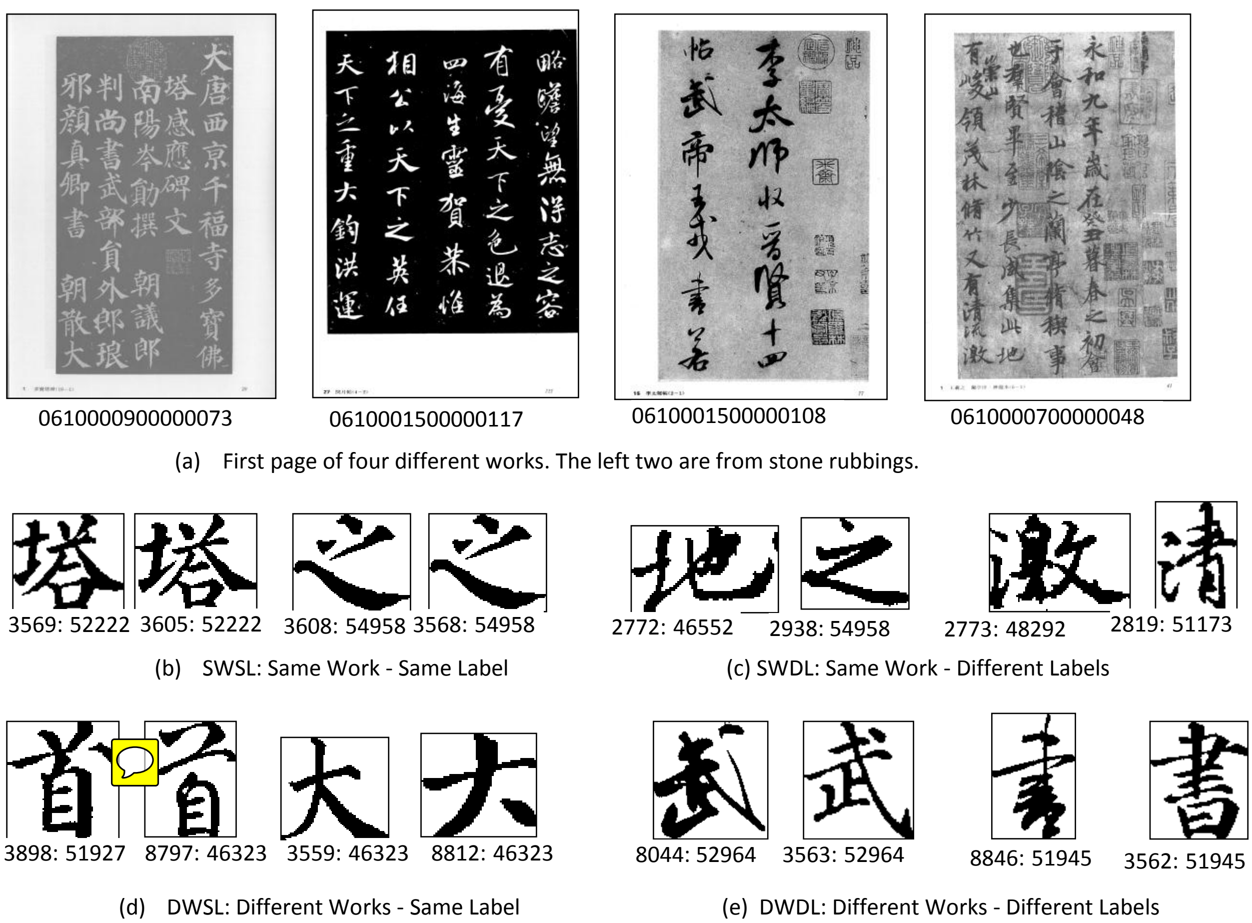


**Figure1. Scanned works and their binarized characters: (a) First row: four scanned first-page images; (b) Pairs same-label images from the same work; (c) Pairs of different labels from the same work; (d) Pairs of same-label images from different works; (e) Different labels from different works. The Character_ID (CID) and the label (GB2312) are shown below each image.**

An analogy with font recognition may clarify the difference between *same-label* and *different-label* classification. In same-label pair classification, the classifier is presented with pairs like SWSL: c-c, g-g; and DWSL: d-d, r-r, to decide whether the first and second letters of each pair are from the same font. Here the fonts are: Bodoni-Bodoni, Bookman-Bookman, and Courier-Arial, Times-Calibri. The classifier would be trained only on same-label pairs from fonts other than these. In different-label pair classification, the pairs may be SWDL: c-d, e-g; and DWDL: d-n, r-y. These letters are drawn from the same fonts as above. The classifier would never have seen any example of any of these typefaces, yet must decide whether each pair of letter is from the same font or not.

We emphasize that our training and test sets are always drawn from different works. This is a realistic but much more stringent constraint than reported in any previous research on calligraphic style. In our experiments, the classifier is not trained on character images drawn from the works whose sources are being compared. We aim to determine whether *differences in style*, such as in stroke-thickness, stroke-uniformity, aspect-ratio, slant, etc., can be characterized independently of specific works and character classes.

In Section 2 we present more precisely the problems to be solved, and the adopted methods of feature extraction and classification. Section 3 describes the experimental design and the experiments conducted. Section 4 summarizes the results of over 250 experiments. In Section 5 we outline related work in progress and suggest how pair dichotomy can be adapted to style identification (rather than style verification) and to forgery detection. Since we don't expect near-perfect algorithmic style identification, for this purpose we hope to adapt an interactive tool, CalliGUI, developed for labeling calligraphic images with the help of an imperfect classifier[11].

## 2. PAIR STYLE CLASSIFICATION

### 2.1 Notation

A character-image I(k) is an $m_k$ x $n_k$ binary array. A unique Character identifier (CID) *k* is assigned to each character image. Although the book pages were scanned to 24-bit RGB format, the characters were subsequently semi-automatically segmented and binarized[3]. The height and width of the characters range from 30 pixels to 450 pixels because different books illustrate calligraphy at different magnifications. We deliberately chose to avoid the inevitable distortion that results from character size normalization and instead use scale-invariant features.

Each character-image *k* has a four-digit Work_ID w(*k*) and a five-digit (16-bit) GB2312 code c(*k*) analogous to ASCII. In the following experiments we do not use the other ancillary information (Book_ID, Page_ID, …) that the database contains for each character image. The feature vector associated with a character is **v**(k), where **v** is a vector of *d* components.

The features extracted from a pair of characters are combined into a *pair-feature* vector $\mathbf{v}(k_i,k_j)$ where each element is the absolute value of the difference between the corresponding elements of the individual feature vectors. Therefore the essential information associated with each character-pair-feature **v**(*k1*,*k2; w1*,*w2; c1*,*c2*) consists of a pair of CIDs *k1*,*k2* that uniquely identify the source images, the identities of source works (Work_IDs *w1*, *w2*) and the GB 2312 source labels *c1* and *c2* of the pair of characters from which the elements of the pair-feature were extracted. The classifier is trained to discriminate between pair-features of the classes C1: *w1=w2* vs. *C2: w1≠w2.* The pair-features fall into the following types (which are unknown to the classifier during both training and testing).

SWSL, where w1=w2, c1=c2, k1≠k2
SWDL, where w1=w2, c1≠c2, k1≠k2
DWSL, where w1≠w2, c1=c2, k1≠k2
DWDL, where w1≠w2, c1≠c2, k1≠k2

The characters selected for each experiment are randomly partitioned into a training set of `N_train` vectors and a test set of `N_test` vectors in such a way that no WorkID can appear in both training and test set. In same-label pair classification, the features represent SWSL and DWSL types. In different-label pair classification, they are SWDL and DWDL types. In either case, the classifier assigns a feature vector into the same-work (SW) or the different-works (DW) class. Figure 2 shows a single example each of the training and test SWSL and DWSL source character pairs for same-label pair classification, and of the corresponding SWDL and DWDL pairs for different-label pairs classification. Style similarities and style differences are less obvious here than in Fig. 1 because, as in our experiments, these samples were selected at random.

When more than one pair of characters is available for style verification, the results can be combined by majority vote of the assigned classes or by multiplying the estimated probabilities of correct classification.

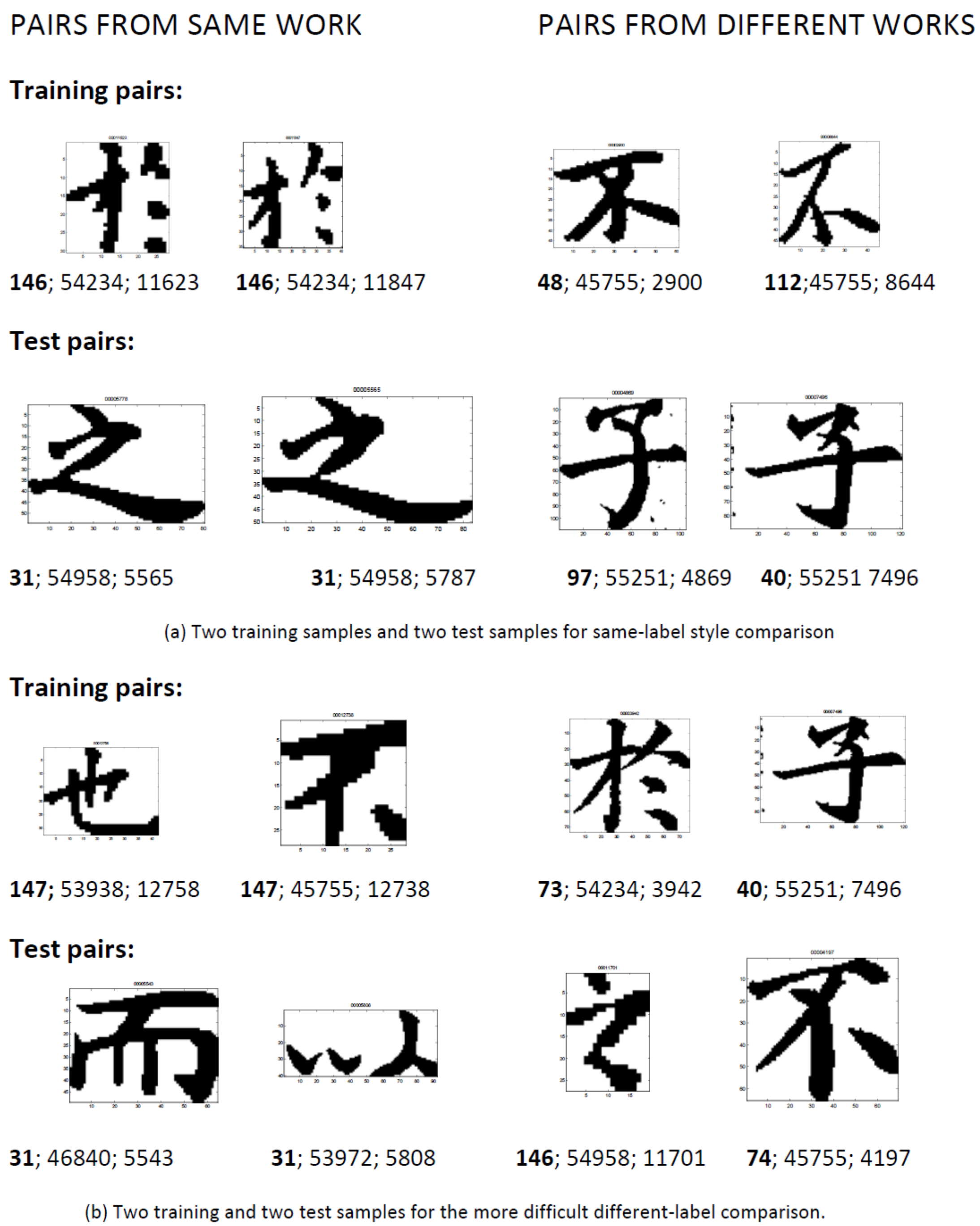


**Figure 2. Each sample consists of a *pair* of either same-label or different-label character images. The work_ID, label and CID are shown below each image.**

### 2.2 Feature Extraction

Almost any set of features, including Exclusive-Or of the bitmaps, should work for same-label pairs classification because most calligraphers pride themselves on the consistency of their brush work. Indeed, some works appear almost as regular as print. The noise introduced by photography (some of the originals are on non-scanner-friendly stone, bamboo sheet, silk scroll and rice paper), digitization, segmentation and binarization is often greater than the variation between the actual characters. It may be possible, at least in principle, to warp same-label images, even from different works, into one another other.

Different-label pair classification is considerably more difficult, because variations in shape define different character classes as well as different styles. We therefore concentrate on simple local features that are more likely to be sensitive to local style variations than to global shape distinctions.

Most of our features consist of normalized counts of the number of distinct locations where various binary template patterns fit into the character bitmap. Examples of such templates are $m \times n$ pixel rectangles of various sizes, aspect ratios, and orientations.

For example, the number of foreground pixels or 1x1 pixel squares (i.e., the area $A$) divided by image height $H \times$ width $W$ is a measure of the density of the character. The aspect ratio H/W is valuable because some calligraphers favor tall characters while others like fat characters. More interestingly, if $Q$ is the count of 2x2 squares of foreground in the image, then $A/(Q-A)$ is a good approximation of average stroke width[12]. Other features target the difference in the thickness of horizontal and vertical strokes, or their average length and slope. Most of these features can be found from the area of the character-image left after morphological erosion with the appropriate kernel. The decomposition of kernels into simple structuring elements leads to rapid feature extraction.

In addition to local features, we use a binary feature to denote whether the character is a stone rubbing (background darker than foreground), and the skewness (third central moment, a measure of asymmetry) of the horizontal and vertical projections of the image.

### 2.3 Classification

We have tried only two simple Bayesian classifiers: a linear classifier with pooled covariance matrix and a quadratic classifier with individual covariance matrices. In most of our experiments the number of training samples barely exceeds the often-recommended minimum of ten times the feature dimensionality. Although other classifiers like Nearest Neighbors or Support Vector Machines may yield higher accuracy, they are not likely to affect conclusions drawn about the relative effectiveness of same-label versus different-label classification, and about the nature of the statistical dependence between same-work images.

## 3. EXPERIMENTS

### 3.1 Experimental Design

The experimental design is complicated by the skewed label frequency distribution of Chinese characters. While there is an over two orders of magnitude difference in letter frequencies in English, the ratio between common and rare characters is much higher in Chinese. The skew is accentuated by the chronological span of the database: not only character shapes, but even character usage (the distribution of label frequencies) have changed considerably through the centuries. The length of the works also varies over a large range. Therefore there is no guarantee that character-images with the same label can be found in two arbitrarily chosen works. Table 1, which shows the co-occurrence of the most frequent labels, gives insight into the limits on the size of our experiments. It is seen, for example, that none of the 15 most frequent labels occur more than once in *both* the first and the third largest work.

For each experiment, we select the `N_works` (typically 170) largest works, and the `N_labels` (typically 800) most frequent labels. The remaining works contain too few characters, and the remaining labels are too rare, to contribute to either training or classification. We normally need only a few images with the same label from the same work, so we keep at most `N_max` (usually 30) occurrences of each character-image with a given label from any work.

**Table 1.Co-occurrence of 15 most common labels in 10 largest works.**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | 30 | 30 | 25 | 30 | 30 | 30 | 30 | 21 | 30 | 27 | 11 | 15 | 23 | 0 | 6 |
| **2** | 30 | 30 | 21 | 28 | 28 | 30 | 30 | 19 | 29 | 26 | 12 | 9 | 18 | 1 | 7 |
| **3** | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 |
| **4** | 13 | 22 | 18 | 8 | 7 | 10 | 1 | 6 | 2 | 7 | 8 | 2 | 12 | 2 | 6 |
| **5** | 12 | 3 | 1 | 12 | 4 | 5 | 5 | 2 | 1 | 2 | 7 | 5 | 0 | 3 | 2 |
| **6** | 10 | 1 | 2 | 4 | 2 | 8 | 1 | 11 | 0 | 2 | 0 | 1 | 1 | 7 | 5 |
| **7** | 9 | 5 | 3 | 2 | 0 | 4 | 3 | 3 | 1 | 0 | 2 | 3 | 0 | 3 | 0 |
| **8** | 2 | 6 | 1 | 0 | 3 | 0 | 0 | 0 | 1 | 6 | 6 | 0 | 0 | 3 | 2 |
| **9** | 1 | 3 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 2 | 3 | 1 | 1 | 1 |
| **10** | 11 | 2 | 1 | 2 | 4 | 0 | 0 | 1 | 2 | 1 | 4 | 3 | 2 | 1 | 0 |

The number of character pairs that participate in voting on whether a sample is from the same or from different works is `N_same`. If `N_same` is 5, then a training or test set of 200 samples would consist of 10 sets of 5 same-work pairs plus 10 sets of 5 different-work pairs. For same-label experiments, this means finding five pairs of works that each have at least 5 samples of a shared label, plus ten works that have at least 5 samples of any label. Different-label experiments impose less restrictive constraints. Part of a test set for different-label pairs classification is shown in Table 2.

**Table 2. Part of a test set. The first six pars are same-work pairs, and the next six are different-works pairs. Here there are three pairs from each work (`N_same`=3).**

| First character | | | | Second character | | | |
|---|---|---|---|---|---|---|---|
| WorkID | Label | CID | SeqNo | WorkID | Label | CID | SeqNo |
| 8 | 54466 | 1776 | 714 | 8 | 54224 | 1780 | 718 |
| 8 | 47557 | 1777 | 715 | 8 | 51645 | 1781 | 719 |
| 8 | 51177 | 1779 | 717 | 8 | 51111 | 1778 | 716 |
| 94 | 54224 | 6063 | 4490 | 94 | 54958 | 6080 | 4506 |
| 94 | 53938 | 6077 | 4504 | 94 | 54234 | 6064 | 4491 |
| 94 | 51906 | 6089 | 4515 | 94 | 50916 | 6074 | 4501 |
| 63 | 53186 | 3350 | 1904 | 131 | 51111 | 10231 | 8421 |
| 63 | 55267 | 3349 | 1903 | 131 | 55242 | 9825 | 8088 |
| 63 | 50414 | 3351 | 1905 | 131 | 47010 | 9689 | 7959 |
| 88 | 46840 | 6423 | 4789 | 36 | 54958 | 7272 | 5634 |
| 88 | 54958 | 6447 | 4813 | 36 | 55251 | 7276 | 5638 |
| 88 | 45755 | 6458 | 4824 | 36 | 46323 | 7280 | 5642 |

The works and labels are selected pseudo-uniformly (without replacement) from the available pools. To select character-images, we take pointers from successive elements of random permutations rather than discrete random numbers. The feature pairs generated from `N_same` image pairs are used to evaluate the increase in accuracy when several sample pairs are available. Each experiment is repeated `N_replicate` times with different seeds for initializing the pseudo-random sequence generator. Different replicates may contain some of the same samples.

Insistence on a truly randomized design precludes using the same test set for all comparisons. Different random selections give significantly different results, and there are not yet enough samples in the database for experiments at the 1% or 5% level of significance. These are not cherry-picked results.

### 3.2 Experiments

Experiments were conducted to determine the effects of:

1. Using same-label vs. different-label pairs of characters to determine whether two multi-pair samples originated in the same or different work(s).
2. Varying the number of training pairs, with a fixed number `N_same` pairs per work.
3. Varying the number `N_same` of test pairs from the same work(s) participating in a vote.
4. Linear versus quadratic classification boundary. The linear classifier used the pooled covariance matrix of all the training samples.
5. Increasing the number of features. The order of the features tested was kept constant, so the best features were always included in the feature set.

Each test set has the same number of same-work and different-works pairs. `N_same`, the number of pairs for each work-label-combination, is the same in the training set and the test set. So if three sample-pairs are used to determine whether two works are the same or different, three same-work and three different-works sample pairs are used in training. The arrangement shown in Table 2 also applies to the training set of that experiment (but with different WorkIDs).

Results from about 200 classification runs, each with a different training set, are reported in Section 4. In all experiments an attempt was made to run eight different replicates with different seeds for the pseudo-random number generator. In some experiments, the size of the data set limited the number of replicates.

An experiment with 200 training pairs and 80 test pairs, replicated eight times, takes a little over a minute on an ancient laptop. Most of the time is taken by reading the selected 560x8 bmp image files from disk.

## 4. EXPERIMENTAL RESULTS

The experiments in Table 3 (8 features, `N_same`=3, `N_test` = 10×3×2 = 60 pairs) show little difference as a function of the size of the training set in either the single-character or the three-votes error rates. There are not enough characters for more than 240 training pairs without reducing the number of test pairs even further. Some errors may be due to multiple works that do not exhibit any observable difference in style because they have the same author. This is likely because there are 211 works but only 50 authors identified in the database.

**Table 3. Effect of training set size (same-label pairs)**

| Training Pairs | % Correct | % Vote Correct | Replicates |
|---|---|---|---|
| 60 | 81 | 83 | 8 |
| 120 | 82 | 81 | 8 |
| 180 | 82 | 84 | 8 |
| 240 | 75 | 80 | 2 |

Table 4 shows that basing the decision on more characters per work is effective. In these experiments the number of pairs in the training sample increases from 20×3×2=120 to 20×17×2=680. The number of test samples is varied from 10×3×2=60 to 10×17×2=340 pairs. For the largest experiment, there were enough samples only for one of the eight replications. The first row of Table 4 is the same as the second row of Table 3.

**Table 4. Effect of number of pairs per work (same-label pairs)**

| Pairs per Work | % Correct | % Vote Correct | Replicates |
|---|---|---|---|
| 3 | 82 | 81 | 8 |
| 5 | 81 | 84 | 7 |
| 7 | 78 | 85 | 5 |
| 9 | 80 | 85 | 4 |
| 11 | 79 | 85 | 1 |

It is much easier to find enough samples without the constraint that the labels of the characters from the works being compared must be the same. Whereas in Table 3 there were enough samples for all eight replications only for 180 training pairs, if we drop the same-label constraint we can train on up to 480 pairs (Table 5). This does not, however, improve the classification results. In Table 5, as in Table 3, there are 5 features, `N_same`=3, `N_test` = 10×3×2=60 pairs.

**Table 5. Effect of training set size (different-label pairs)**

| Training Pairs | % Correct | % Vote Correct | Replicates |
|---|---|---|---|
| 60 | 78 | 80 | 8 |
| 120 | 79 | 80 | 8 |
| 180 | 78 | 80 | 8 |
| 240 | 78 | 80 | 8 |
| 300 | 77 | 79 | 8 |
| 360 | 77 | 79 | 8 |
| 420 | 77 | 79 | 8 |
| 480 | 76 | 79 | 8 |

In Table 6 as in Table 4 there are 8 features, `N_train` increases from 120 to 680, and `N_test` from 60 to 340. Increasing the number of pairs participating in the vote does not help because results from the same works are highly correlated: either most of a set of pairs are classified correctly, or almost none are. Nevertheless the classification accuracy in the different-label experiments of Tables 5 and 6 is not far from that of the corresponding same-label experiments of Tables 3 and 4, justifying our claim that the features used are almost as sensitive to style as to individual character shape.

**Table 6. Effect of number of pairs per work (different-label pairs)**

| Pairs per Work | % Correct | % Vote Correct | Replicates |
|---|---|---|---|
| 3 | 79 | 80 | 8 |
| 5 | 80 | 80 | 8 |
| 7 | 78 | 80 | 8 |
| 9 | 80 | 81 | 8 |
| 11 | 80 | 80 | 8 |
| 13 | 79 | 79 | 8 |
| 15 | 77 | 79 | 8 |
| 17 | 79 | 79 | 8 |

Even in our largest experiments, there are not enough samples for good estimates of the class-conditional covariance matrices necessary for quadratic classification. In the same-label experiment of Table 7, there were 120 training pairs, 60 test pairs, 5 votes, and 8 features. This experiment was run with only five features because the SWSL half of the training set has a singular covariance matrix because the foreground color is always the same in both members of a pair.

**Table 7. Quadratic vs. linear classifier**

| Classifier | % Correct | % Vote Correct |
|---|---|---|
| Linear | 73 | 78 |
| Quadratic | 72 | 75 |

Table 8 shows that we have not yet found enough good style features to add to our eight basic features. The other parameters are the same as in Table 7.

**Table 8. Feature dimensionality**

| Number of features | % Correct | % Vote Correct |
|---|---|---|
| 5 | 73 | 78 |
| 10 | 80 | 84 |
| 15 | 77 | 83 |

## 5. DISCUSSION

We reported a set of baseline experiments on calligraphic style discrimination using the CADAL database. The objective of these experiments was to determine whether pairs of samples originate in the same work or in two different works, without the classifier being trained on either work. The results show the accuracy attainable with eight simple features and a simple off-the-shelf classifier. Even with improved features and more data, classification accuracy will fall short of 100%. Any practical application will require an effective interface for human interaction.

As expected, it is easier to tell whether the same calligrapher created two characters with the same label than whether he or she created two randomly-selected characters with different labels. Using several pairs for each decision yielded only small improvements in accuracy in either case. In our experiments same-label classification plateaued at ~85% and different-label classification at 80%. The small difference suggests that our features are sensitive to *style differences* that are independent of the shape characteristics that determine the label of a character.

The somewhat complex experimental protocol is a plausible template for experimentation on style-specific shape features. Its randomized design provides some safeguards against over-reaching conclusions. It may challenge some mindsets about the nature of style differences.

This research is also intended to open the way to style identification. If scholars tag the works in a calligraphic data base with style labels, then a style can be assigned to a work with unknown style by determining its style-similarity with each of the tagged works. While this *n*-class problem can be approached through a series of dichotomies, that is not likely to be the most efficient method.

Finally, we hope that style discrimination will provide insight into the detection of calligraphic forgeries. In many practical situations, however, physical examination – chemical assays, spectrography, x-rays, carbon-dating – may well prove more conclusive than any conceivable digital version of the traditional document examiner's loupe. In this context, forgery detection is complicated by the need to distinguish between copies made by admirers and students of a master calligrapher, and copies meant to deceive in the expectation of financial gain.

## ACKNOWLEDGMENT

We thank RPI student Ashfaqul Islam for improving CalliGUI and gratefully acknowledge the support of the Rensselaer Center for Open Software and of the Science & Technology Program of Shanghai Maritime University

## REFERENCES

[1] CADAL web site: http://www.cadal.zju.edu.cn

[2] Universal Digital Library web site: http://www.ulib.org

[3] Zhang, X., Nagy, G., The CADAL Calligraphic Database, Procs. Workshop on Historical Document Imaging and Prcessing (HIP 2011), Beijing, 2011.

[4] Zramdini, A.W., Ingold, R., Optical Font Recognition Using Typographical Features. IEEE Trans. Pattern Anal. Mach. Intell. 20(8): 877-882,1998.

[5] Gregory R. Ball, G.R. Pu, D., Stritmatter, R., Srihari, S.N., Comparison of historical documents for writership. Procs. Document Recogniton and Retrieval, DRR 2010.

[6] Jain, A.K., Ross, A. and Flynn, P.J., (eds.), Handbook of Biometrics, Springer 2007.

[7] Srihari, S.N., Computational Methods for Handwritten Questioned Document Examination, Final Report, Award Number: 2004-IJ-CX-K050, U.S. Department of Justice, http ://www.ncjrs.gov/pdffiles1/nij/grants/232745.pdf

[8] Choi, S., Cha, S., Tappert, C.C., A Survey of Binary Similarity and Distance Measures, Journal of Systemics, Cybernetics and Informatics, Vol 8 No 1 2010, pp 43-48.

[9] Zhang, X., Zhuang, Y., Visual Verification of Historical Chinese Calligraphy Works, Lecture Notes in Computer Science, MMM'2007, LNCS 4351, pp: 354–363, 2007.

[10] Zhuang, Y., Lu, W., Wu, J., Latent Style Model: Discovering writing styles for calligraphy works. J. Visual Communication and Image Representation 20(2): 84-96, 2009.

[11] Nagy, G., Zhang, X., CalliGUI: Interactive Labeling of Calligraphic Character Images, Procs. ICDAR 2011, Beijing, 2011.

[12] Bartz, M.R., The IBM 1975 Optial Page Reader Part II: Video Thresholding System, IBM Journal of Research and Development 12:5, pp. 354-363, 1968.