View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Institutional Knowledge at Singapore Management University

Singapore Management University
Institutional Knowledge at Singapore Management University

Research Collection School Of Information Systems | School of Information Systems

2-2009

Stochastic Modeling Western Paintings for Effective Classification

Jialie SHEN
Singapore Management University, jlshen@smu.edu.sg

DOI: https://doi.org/10.1016/j.patcog.2008.04.016

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research

Part of the Databases and Information Systems Commons, and the Numerical Analysis and Scientific Computing Commons

Citation

SHEN, Jialie. Stochastic Modeling Western Paintings for Effective Classification. (2009). *Pattern Recognition*. 42, (2), 293-301. Research Collection School Of Information Systems.
Available at: https://ink.library.smu.edu.sg/sis_research/758

This Journal Article is brought to you for free and open access by the School of Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Published in Pattern Recognition, 2009 Feb, Volume 42, Issue 2, pp. 293–301.
http://doi.org/10.1016/j.patcog.2008.04.016

Stochastic modeling western paintings for effective classification

Jialie Shen*

School of Information Systems, Singapore Management University, 80 Stamford Road, Singapore 178902, Singapore

A B S T R A C T

As one of the most important cultural heritages, classical western paintings have always played a special role in human live and been applied for many different purposes. While image classification is the subject of a plethora of related publications, relatively little attention has been paid to automatic categorization of western classical paintings which could be a key technique of modern digital library, museums and art galleries. This paper studies automatic classification on large western painting image collection. We propose a novel framework to support automatic classification on large western painting image collections. With this framework, multiple visual features can be integrated effectively to improve the accuracy of identification process significantly. We also evaluate our method and its competitors based on a large image collection. A careful study on the empirical results indicates the approach enjoys great superiority over the state-of-the-art approaches in different aspects.

Keywords:
Classic western paintings
Identification
Image retrieval
Cultural heritage

1. Introduction

In recent years, the growth of the Internet has unleashed a wave of innovations reshaping the way people interacting with digital media data (e.g. music, image and video). Massive image data is being generated on a daily basis from various domain applications [1–8]. In particular, thousands of paintings have been collected by museums and galleries for the purpose of exhibition, historical study and education. Consequently, developing effective technologies to organize, manage and access those visual information has been becoming an extremely important research issue. In particular, automatic image classification has been received a lots of attention in the past decades. With such techniques, richer functionality for browsing and exploring can be possible for general users.

While image classification is the subject of a plethora of related publications, relatively little attention has been paid to automatic categorization/classification of western classical paintings which could be a key technique of modern digital library and related fields. In addition, due to huge amount potential applications including indexing, retrieval and online recommendation systems, techniques facilitating classification on large classical western painting collections are gaining in importance. The classical western paintings date back to the Ancient Egypt, about 6000 years ago. They are one of the most important components for the Human cultural heritage. In this study, image classification problem is explored for western paints during Baroque and Rococo period [9,10]. This period began around 1600 B.C. and continued throughout the 18th century.

The image classification problem is to determine predefined categories for an unlabeled image based on its computable content parameters. Traditionally, the categorization process can be divided into two main steps: feature extraction and multiclass classification. In the step of feature extraction, image analysis methods are used to generate representation for visual content. Then, a classifier, which can be an algorithm or a statistic model, is applied to identify the labels from the representation of the images with respect to their features. The obvious shortcoming of the approach is that it requires a comprehensive image signature to represent visual content well. However, it has so far proved that using low level features to represent high-level concepts ("semantic relevant") effectively is a highly difficult task due to the well known "semantic gap". In fact, recent research results in psychophysics and neurophysiology have shown that different kinds of global and local features are very important for visual information classification [11,12]. Thus, minimizing the "semantic gap" in the respect of image content representation introduces two sub-problems:

- How should the low-level features be combined effectively for a given classification or retrieval tasks, and
- How should the features extracted in different solution be combined for classification or retrieval tasks.

Furthermore, recent research results revealed that: for categorization task, it can lead to various problems such as poor robustness and

* Tel.: +65 68280982.
E-mail address: jlshen@smu.edu.sg.

Fig. 1. Examples of Claude Monet's paintings.

small sample size problem by only utilizing a single source of visual characteristic from image.

It is not hard to see that classic western paintings drafted by the same artist share certain drawing characteristics. They could contain similar visual patterns and styles. For particular painters, the number of style can be very limited. The typical example is the founder of impressionist painting—Claude Monet [13]. His paintings include visible brushstrokes, open composition, emphasis on light in its changing qualities (often accentuating the effects of the passage of time), ordinary subject matter, the inclusion of movement as a crucial element of human perception and experience, and unusual visual angles. Fig. 1 illustrates some examples of Claude Monet's paintings. All of those observations suggest the feasibility of automatic categorization process based on visual information automatically. In this study, a general solution for classical paints image categorization has been developed and we have carried out a relative performance evaluation over a large scale image collection. The goal of this research is to develop a approach that enables a computer to examine a piece of classical paint, and recognize who is the artist to draw the paint effectively. To the best of our knowledge, no similar work has been reported in the literature. In brief, the main contributions to this advancement and content of the article are as follows:

- We propose a novel statistical model to profile salient characteristics of art work from the certain painter. The approach can effectively combine multiple kinds of visual features at different resolutions.
- Distinguished from previous approaches, we design a hieratical structure that integrates different classification models seamlessly. To further improve accuracy, multilevel score fusion scheme is also designed based on regression model. A prototype implementation of the proposed scheme has been implemented with Java and C++.
- To validate the superiority of the proposed method and provide comparative results, we have constructed a large test collection. This testbed is the first large image data set for the classical painting classification problem. It contains 1080 classical painting images from 25 different artists. Based on this test collection, we have carried out a comprehensive empirical study. The careful analysis of the results indicates that our approach achieves substantial performance improvement on different aspects over the existing methods.

The remainder of this paper is organized as follows. The next section reviews some background information and related work about painting image classification. Section 3 presents the information about local and global feature extraction. Section 4 describes the detailed algorithms of the proposed scheme. Section 5 gives an introduction about the experimental configuration. In Section 6, we present a performance study over the proposed approach and give a detailed analysis of results. Next, issues related to system architecture and configuration are discussed in Section 7. Finally, we draw some conclusions and indicate future directions for this research in Section 8.

2. Related work

The main goal for image classification is to assign the image to a predefined class/label. The basic process can be used to speed up image retrieval and the effectiveness of visual information management processing. Driven by various emerging domain applications, a large amount of innovations have been accomplished for classification in the recent years. The corresponding research study to overcome obstacles to accurate image classification is directly associated to several technical fields, among which image feature and content representation extraction, classification and visual information modeling with statistical approaches and performance evaluation. Due to space limitation, it is possible to give a comprehensive survey covering related research. Instead, we drill down our focus and provide some research results directly associated to this study. Details about basic principles and a more comprehensive study can be found in Ref. [2].

Although a lot of research efforts have been invested in the domains of image classification/cagetoerization, very limited attention has been focused on developing techniques relate to accident paintings. Among the earliest of such systems, Melzer et al. explored how to apply different type of brush strokes to classify portrait miniatures [14]. Both a model based and a semiparametric, neural network approach are used as the detector. In the following work, Sablatnig et al. [15] study using "structural signature" based on brush strokes for artist classification problem. A hierarchically structured classification scheme is developed to separates the classification into multiple kinds of information: color, shape of region, and structure of brush strokes. In addition, the problem of painting image retrieval problem has been investigated by Colombo et al. They discuss the emotional effects of different visual features [16].

Li and Wang develop a general classification framework for Chinese painting classification problem [17]. They design a mixture of 2D multiresolution hidden Markov models (MHMMs) as statistical modeling method to characterize a type of stroke from different artists. After training the models, the methods can be applied for categorization of images. In the study, the Daubechies 4 wavelet coefficients are considered as the content representation for each image. The algorithm has been evaluated using test collection containing 271 paintings from five different Chinese artists. The best accuracy achieved is around 80% dependent on the painters. Authors also claims that the approach can be potentially applied to classification of other cultures. However, they did not provide any corresponding experimental results. In Ref. [18], Jiang et al. proposed a scheme to categorize traditional Chinese paintings into two classes—Gongbi and Xieyi. Low-level features considered include color histogram, color coherence vectors, texture features and edge size histogram. In the system, C4.5 decision tree classifier is firstly

Table 1
Summary of existing identification methods' properties

Identification methods	Multifeature integration	Size of testbed	Local feature considered	Global feature considered
MKZ [14]	No	Small	Yes	No
CDP [16]	No	Small	Yes	No
LW [17]	No	Small	No	No
JHY [18]	No	Small	No	No

used as a pre-classification to identify traditional Chinese paintings and non-traditional Chinese paintings. Then SVM (support vector machine) is applied as the final classifier. They report an accuracy of 91.01% based on color coherence vector and 90.3% based on Ohta color histogram. The corresponding test collection contains 3688 images. More recently, Berezhnoy et al. initializes the AUTHENTIC project [19]. The main aim is to create a set of software tools to help artists in assessing authenticity of paintings.

Based on above discussion and analysis, Table 1 summarizes the properties of the previous approaches for classification of identification methods. Those methods either use single kind of visual feature to represent painting objects or simply ignore the joint effects of different feature. Furthermore, none of them considers both local and global features. In this case, achieving effective visual information retrieval is rather difficult. Moreover, all of methods mentioned above have been tested based on very small data sets.

3. Visual feature extraction

Local and global visual information could play different roles in image classification and retrieval. Both of them contain various discriminative evidences related to image identification. Thus, our proposed system considers both of them. In this section, an introduction about visual feature and related extraction procedure is given. In our system, those features are used as content representation for images.

3.1. Global visual features

Global visual features contain image characteristic from the coarser level [20]. For each image, our approach extracts our different global features and their detail information is present as below:

- *Color*: It is known that the human eye responds well to color feature. Given a discrete color space defined by some color axes, the color histogram is obtained by discretizing the image colors and counting the number of times each discrete color occurs in the image. In our experiments, we used the color space CIE $L*u*v$. The reason for selecting the CIE $L*u*v$ instead of the normal RGB or other color spaces is that it is more uniform perceptually [21]. The three axes of the $L*u*v$ space are firstly divided into 4 sections to obtain a total of 64 (i.e. $4\times4\times4$) bins for the color histogram. However, we found that, for the collection of images used in our experiments, not all the bins had non-zero counts. So, after eliminating those bins which had zero count, our color features are presented as 37-dimensional vectors.
- *Texture:* Texture features carry the property measures, such as smoothness, coarseness and regularity, of an image. The texture features are extracted using a filter-based method. This method detects the global periodicity of intensity values in an image by identifying regions that have high energy and narrow peaks. The advantage of filter-based methods is in their consistent interpretation of feature data over both natural and artificial images. The Gabor filter [21] is a frequently used filter in texture extraction. It measures a set of selected orientations and spatial frequencies. The total number of filters needed for our Gabor filter is 30. Texture features are therefore represented as 30-dimensional vectors.
- *Shape:* Shape is an important and powerful attribute for image retrieval. It can represent spatial information that is not present in color and texture histogram. In our system the shape information of an image is described based on its edges. A histogram of the edge directions is used to represent global information of shape attribute for each image. The Canny edge operator is used to generate edge histogram of images in the pre-processing stage. In order to solve the scale invariance problem, the histograms are normalized to the number of edge points in each image. In addition, smoothing procedures presented in are used to make histogram invariant to rotation. The histogram of edge directions is represented by 30 bins. Shape features are thus presented as 30-dimensional vectors.
- *Color layout:* A color histograms describes the global color distribution in a image [22]. However, it lacks information regarding the spatial distribution of color. Color layout is used to capture local information on the color distribution. The whole image is firstly divided into 8×8 sub-blocks and their average colors are calculated. A series of coefficients are obtained by conducting 8×8 DCT (Discrete Cosine Transform) over these average colors. Low-frequency coefficients are selected using zigzag scanning and quantized to construct this color layout descriptor. The color space for this descriptor is YCrCb. The dimensionality of color layout is 30. This includes the first 10 value from Y, Cr and Cb coefficients.

3.2. Local visual features

As mentioned above, our approach considers both local and global visual features. To extract local features, special scheme is designed to integrate joint effect of different partitions from an image. The corresponding extraction process can be denoted as

$$v_f = Extract_f(obj) = [v_{1f}, v_{2f}, \ldots, v_{Bf}] \tag{1}$$

where v_f is the set of vectors for feature type f extracted from the B blocks of the input image object obj. One 2D image example is illustrated in Fig. 2. In this study, image is partitioned into $4\times4=16$ subblocks and 16 Gabor wavelet based feature vectors are extracted from the individual subimage. Thus, B is equal to 16 for this research. Based on Ref. [23], 2D Gabor wavelets can be defined as follows:

$$\psi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{-\|k_{u,v}\|^2\|z\|^2/2\sigma^2}[e^{ik_{u,v}z} - e^{-\sigma^2/2}] \tag{2}$$

where $k_{u,v} = k_v e^{i\phi_u}$; $k_v = k_{max}/f^v$ gives the frequency, $\phi_u = u\pi/8$, $\phi_u \in [0,\pi)$ provides the orientation. It is not hard to find out that Gabor wavelet is made up of a planar sinusoid multiplied by a two dimensional Gaussian. This property insures that the frequency information close to the center of the Gaussian can be easily captured. Since Gabor wavelet can effectively exhibit characteristics of spatial locality and orientation selectivity, image descriptor based it can carry more comprehensive local information about paintings style.

4. A framework for classical western paintings classification

As the architecture shown in Fig. 3, our classification framework contain two major layers—RBF neural network classifiers based on different local and global visual features and classification score fusion schemes based on linear regression [24].

4.1. RBF neural network based classifier

For the purpose of effective classic painting identification, RBF neural network classifier is constructed with multiple kinds of

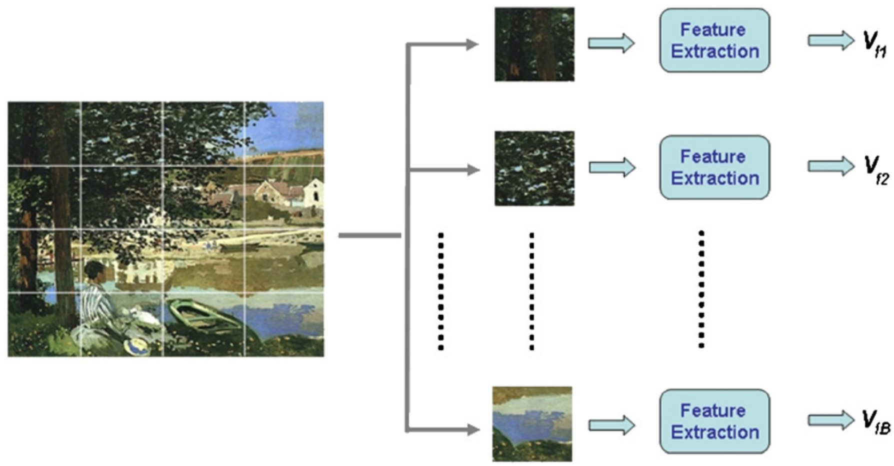

Fig. 2. The procedure of local visual feature extraction.

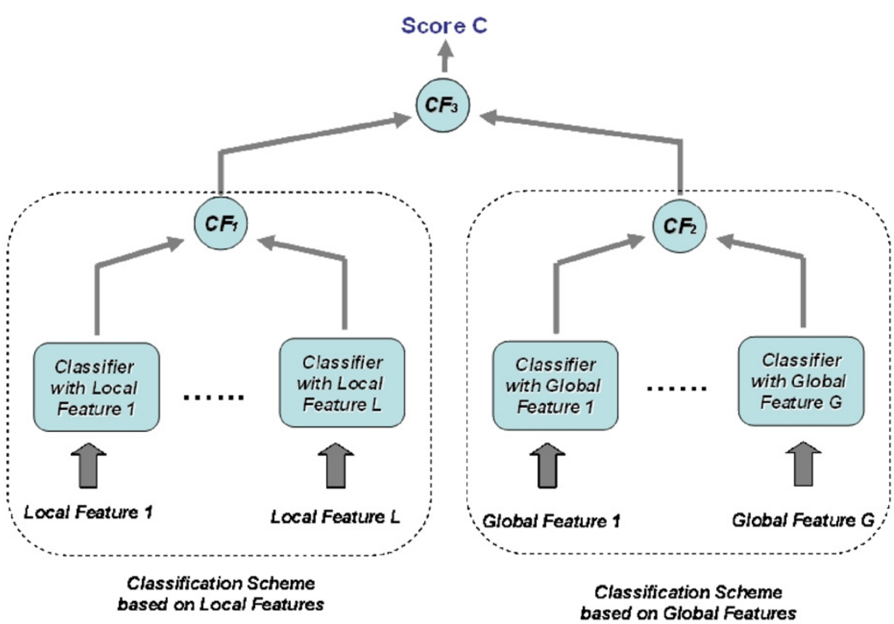

Fig. 3. System architecture of western paintings classification framework.

features. The radial basis function (RBF) neural network is very powerful scheme for data modeling and has been widely applied for function approximation, pattern classification, data compression and so on [25].

In our framework, the individual feature of the visual signal is extracted, and then, individual classifier is built up based on each kind of feature. The structure of the RBF neural network in our framework is shown in Fig. 4 and contains three layers including the input layer, the hidden layer and the output layer. For RBF neural network, RBF represent a special class of functions. Their particular feature is that the functions response monotonically decreases (increase) as the distance from the central point grows. Given the input feature vector V_f, the typical example used in our framework is Gaussian function as

$$R_k(V_f) = \exp \frac{\|V_f - \mu_k\|^2}{2\sigma_k^2} \tag{3}$$

where σ_k represents the standard deviation of the Gaussian function on neuron k at hidden layer. The output layer is a limit discriminant that outputs a weighted sum of the basis functions and the equation for a single output o_c^f from neuron in output layer is as

$$o_c^f(V_f) = \sum_{k=1}^{K} w_{ck} R_k(V_f) + w_{0k} \tag{4}$$

where w_{0k} is the weight of the bias and distance between output vector $O^f = \{o_1, o_2, \ldots, o_C\}$ and standard training pattern encoded is used to quantify similarity between class label and input object using feature vector v_f of feature f as content representation for the input image.

4.2. Training process with binary output codes with penalty

To improve the effectiveness of our categorization model under different circumstances, an encoding scheme called binary output codes with penalty (BOCP) is designed and integrated into our system. It is based on error correcting output codes (ECOC) [26]. The basis of this method is to create codeword for each class and the

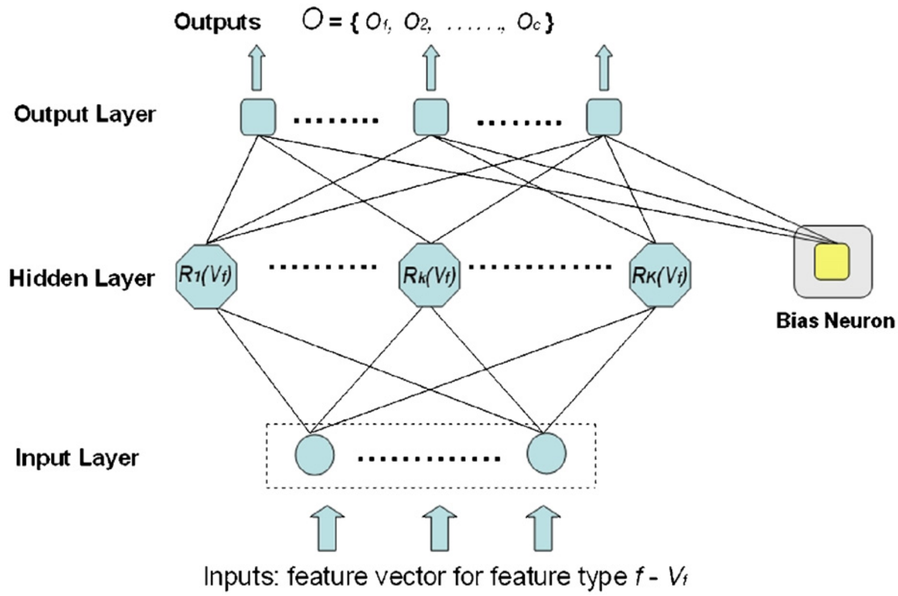

Fig. 4. The structure of radial basis function (RBF) neural network classifier.

determine class label based on distances between outputs from classifier and predefined codewords. Hamming distance is one of the most popular function [27] and used in this work. For the proposed BOCP, the fundamental principle consists of two steps:

- *Training stage*: The first step is to construct $C \times C$ confusion matrix *CM* and the format is as

$$CM = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1C} \\ p_{21} & p_{22} & \cdots & p_{2C} \\ \cdot & \cdot & \cdots & \cdot \\ p_{C1} & p_{C2} & \cdots & p_{CC} \end{bmatrix}$$

  where C is the number of artist in target database and each element p_{ij} represents the probability of painter i's test data but misclassified to painter j by a set of RBF neural networks during training. Then, based on the content of the confusion matrix M, the target vector corresponding to painter c, $TC_c = [tc_{c1}, \ldots, tc_{cC}]$, can be obtained via the following encoding scheme, where tc_{ij} is designed as

$$tc_{ij} = \begin{cases} 1 & \text{if } p_{ij} > 0, \ \ i = j; \\ -1 & \text{if } p_{ij} > 0, \ \ i! = j; \\ 0 & \text{otherwise} \end{cases}$$

  In fact, the target vector in such encoding scheme guarantees that an unique target codeword $TC_c = [tc_{1C}, \ldots, tc_{cC}]$ corresponds to each artist c.
- *Test stage*: The pre-trained RBF neural network produces an outcome, $O^f = [o_1^f, o_2^f \ldots o_C^f]$, based on input feature vector v_f. The distance between output vector O and each target codeword for each painter $TC_c = [tc_{c1}, \ldots, tc_{cC}]$ is calculated with Hamming distance as

$$d_f^c(O, TC_c) = l_f^c = \sum_{c=1}^{C} |o_c^f - tc_{cC}| \tag{5}$$

  The distance $d_f^c(.)$ can be used to measure likelihood value l_f^c between artist label c and input object based on feature type f.

From above, it is not hard to find out that $TC_c = [tc_{1C}, \ldots, tc_{cC}]$ is a desirable output pattern and can be used for the network training patterns.

4.3. Combining classification score and classification process

One obvious question is what is the relative weight of information carried by different kinds of features in a given image data set. As we mentioned previously, different sets of features have different importance to determine the final likelihood score of image. As shown in Fig. 3, the fusion scheme can be viewed as two layers of gating networks and gating function $CF(.)$ is used to combine outputs from RBF neural network classifiers based on different features or gating function for fusion in different layers can be defined as

$$L_l^c = CF(W_l^c, L_{l-1}^c) \tag{6}$$

where $CF(.)$ is combination function, W_l^c is weight vector for combination. This is a multiple layer architecture in the proposed framework. Input of the layer one is output vector containing similarity scores generated by different RBF neural network classifiers using various features. The input of the layer two are outputs from gating function in the first layer. In this study, we choose linear regression as gating function [28]. It has been widely used in the statistical, database and machine learning community. As mentioned before, global features and local features have different impacts on painting categorization process. Global visual information capture the characteristic of paintings in general granularity and at the same time, local features carry more detail property in the finer representation. The combination of two kinds of characteristics can help to enhance final performance of whole system.

Algorithm for automatic painter identification
Input: A painting image with unknown artist c
Output: Painter label of the image
1 Process the image to get segments having fixed size;
2 Feature extraction;
3 Derive relevance scores using architecture shown in Fig. 3;
4 Return artist label with Eq. 7;

After the setup of the system, we can conduct the task of classification. The basic procedure is shown in Algorithm 1 and contains four steps. For a given image item, at the initial stage of the process, the system partitions the small segments (line 1). After that, the feature extraction procedure generates four different kinds of features with the same procedure described before. Next, the features are fed into RBF neural network classifiers. The likelihood score based on a particular feature can be generated based on Eq. (5). Then, relevance scores can be combined with fusion method mentioned above. Those scores quantify the similarity between the input image and the artist label. In the final step, a label for the image can be assigned based on those relevance scores.

$$c^* = \underset{1 \geqslant c \geqslant C}{\operatorname{argmax}} L^C \tag{7}$$

5. Experimental configuration

We now proceed to describe the experimental setting for evaluating the classification framework described in Section 4. It includes test data sets for the experiments. We also present performance evaluation metrics.

5.1. Test collection

In this study, we construct one data set for the our experimental study and the raw data is collected from Internet. It consists of 1080 images from 25 different classical painters. The artists includes Michelangelo Merisi da Caravaggio, Peter Paul Rubens, Jan Vermeer, Diego Velazquez, Nicolas Poussin, Giovanni Battista Tiepolo, Antoine Watteau, Jean-Honoré Fragonard, François Boucher, Thomas Gainsborough, John Constable, Jean Auguste Dominique Ingres, Eugène Delacroix, J.M.W. Turner, Caspar David Friedrich, Francisco Goya, Jacques-Louis David, Claude Monet, Albert Bierstadt, Jean-Baptiste-Camille Corot, Gustave Courbet, Jean-Baptiste-Siméon Chardin, Vincent van Gogh, Edgar Degas and Diego Velázquez. Those artist are active painters from the 16th to the 18th century. The detail information about this data set can be found in Table 2.

The size of each image item in the test collection data set is either 512×256 or 256×512. For classification, we use 20% of each data set for training purposes and the remaining data items to evaluate the performance of all the schemes studied. In all the experiments presented below, there is no overlap between the training set and the test set.

Table 2
Detail information of the classical paintings test collection

No	Artist name	Number of paints
1	Michelangelo Merisi da Caravaggio	40
2	Peter Paul Rubens	50
3	Jan Vermeer	80
4	Diego Velazquez	70
5	Nicolas Poussin	60
6	Giovanni Battista Tiepolo	30
7	Antoine Watteau	50
8	Jean-Honoré Fragonard	60
9	François Boucher	30
10	Thomas Gainsborough	40
11	John Constable	40
12	Jean Auguste Dominique Ingres	50
13	Eugène Delacroix	20
14	J. M. W. Turner	40
15	Caspar David Friedrich	30
16	Francisco Goya	80
17	Jacques-Louis David	40
18	Claude Monet	50
19	Albert Bierstadt	25
20	Jean-Baptiste-Camille Corot	35
21	Gustave CourbetHaydn	40
22	Vincent van Gogh	30
23	Edgar Degas	20
24	Jean-Baptiste-Siméon Chardin	20
25	Diego Velázquez	20

5.2. Evaluation metrics

There are many different metrics for classification performance assessment. To evaluate the effectiveness, we use categorization accuracy (*CA*) for performance assessment in this study and its definition is given as

$$CA = \frac{\textit{Number of objects identified correctly}}{\textit{Total number of input objects}} \tag{8}$$

In order to study the efficiency of the methods, the average response time for processing is applied to measure the efficiency of different techniques,

$$\textit{Average response time} = \frac{\textit{Total query response time}}{\textit{Number of queries}} \tag{9}$$

Where *Total query response time* is the time required for the system to identify the set of images and its total size is equal to *number of queries*. It represents the average time required for classifying a single input image. A lower *average response time* is preferred as it implies faster identification process, and hence better query efficiency.

6. Experimental evaluation

In this study, we focus on a comprehensive empirical study of this framework from three different aspects—effectiveness, efficiency and robustness.

6.1. Effectiveness study

We proceed to make a comparative study on the accuracy of the various artist identification. Table 3 illustrates the results based on test set mentioned in Section 5.1. The bottom two rows show how the CDP and MKZ perform. The experiments verify the fact that both of them achieve very close identification accuracy. CDP's accuracy is 40.2% and MKZ achieves the accuracy of 37.7%. Both are the worst among all five methods. Furthermore, although the JHY technique achieves better performance than CDP and MKZ, it still suffers from low accuracy. This is because JHY only considers global low-level features (such as color histogram, color coherence vectors, texture features and edge size histogram) in one image. At the same time, comparing to previous three methods, LW achieves further improvement. However, performance gain is very limited and the accuracy is still far below 55%. In fact, experimental results clearly demonstrate that method we proposed significantly outperforms the other four approaches. For example, from Table 3, it shows that compared to LW, the SHEN method improves the identification precision from 52.7% to 69.7%. In addition, this is a remarkable improvement over LW. On average, around 32.3% improvement can be observed. On the other

Table 3
Query accuracy comparison of different approaches for artist identification

Query methods	Identification accuracy (%)	Identification accuracy$^+$ (%)
SHEN	69.7	69.6
LW	52.7	52.5
JHY	47.5	47.5
CDP	40.2	40.1
MKZ	37.7	37.4

SHEN denotes the method present in this paper. Identification Accuracy$^+$ denotes results obtained with 5-fold cross validation.

hand, to enhance the stability and robustness of the empirical study result, we also validate the approach with *K*-fold cross-validation. *K* as set to be 5. The resulted achieved is very similar when compared each other. We believe that this performance gain is due to local and global visual information integration.

Table 4
Average response time comparison of different approaches

Query methods	Ave. response time (s)
SHEN	0.29
LW	0.42
JHY	0.47
CDP	0.49
MKZ	0.51

SHEN denotes the method present in this paper.

6.2. Efficiency study

For large visual information systems, size of collection could be big and thus the response time is another crucial concern of the system performance. In this experiment, we show how this affects the time efficiency. Tests were run with 800 images covering

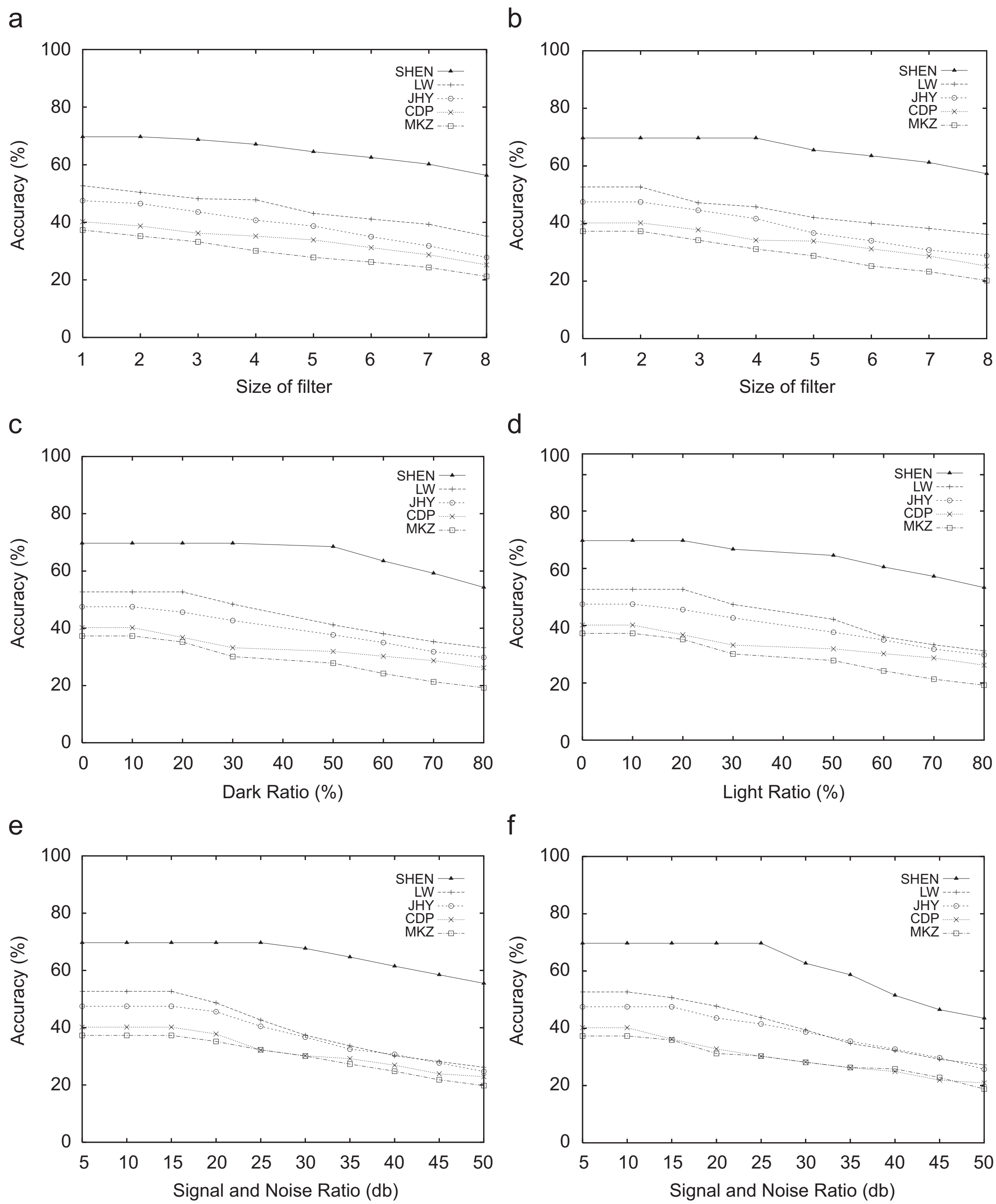

Fig. 5. Comparison of identification accuracy for different categorization methods under various kinds of visual distortion and noise: (a) blurring with Gaussian filter; (b) blurring with median filter; (c) dark; (d) brighten; (e) Gaussian noise; and (f) salt and pepper noise.

all artists in the collection. There is no overlapping between query sets and training sets.

Table 4 shows the response time of query for different artist identification schemes. From the experimental results summarized in the table, we can see that our method leads to the best result in terms of query response time against the other four methods. For example, the artist identification process with LW, JHY, CDP and MKZ required 0.42, 0.47, 0.49 and 0.51 s. Actually they are very close to each other. In contrast, it takes SHEN only 0.29 s to complete the whole process. Comparing other four approaches, there is around 45% saving on average. This is a big improvement.

6.3. Robustness study

Robustness against distortion or noise is one of the most desirable performance factor for large visual information systems. In general, humans have a very impressive capability to identify and classify image or other visual information, from a very small sample and even in the presence of moderate amounts of distortion. This property is potentially useful in various real world application, where the query image may have its origins in a process like low-quality live recording and additional noise in production process. In following, we will study the problem in detail.

To study the robustness of different identification techniques, image data items we used are modified with different kinds of distortion as query examples and a series of experiments have been carried out to evaluate the performance of different systems in the presence of moderate amount of noise and other kinds of distortion. During the evaluation, we ran the same set of tests as in the previous experiments. However, each query image is distorted and the results are compared against the results obtained from using a non-distorted query. The levels and types of distortions are varied to test the robustness of different schemes. Fig. 5 summaries the accuracy of different methods under various distortions. Distortion and noise cases include blurring with Median and Gaussian filter, brighten, darken, Gaussian and salt and pepper noise. The experimental results clearly show that compared with the other four approaches, the SHEN emerges as the most robust technique in terms of accuracy. It performs significantly better than the other competitors on all distortion cases. Thus, we can conclude that SHEN is fairly robust to different levels of noise and distortion. The main reason for this superior performance is that the proposed method considers more information from different aspects and different resolutions. Those information make the method proposed more robust against different distortion and noise cases.

7. Discussion

In this section, we give a discussion of the issues related to the performance on the proposed method. They include: (1) effects of different feature configurations and (2) effects of parameters for RBF neural network. The results present below is obtained with image data sets described in Section 5.1.

7.1. Effects of different feature configurations

Local and global visual information can make different contributions on the painting classification process. The empirical study present in this section tries to answer two questions—whether different kinds of visual features can influence the effectiveness of our approach and if so, how they have an impact on identification accuracy. To study the problems in details, a series of experiments have been designed and carried out, which incorporated different kinds of visual features into our system separately and the results of different system setting based on a single set of queries are compared.

Table 5
Performance comparison of our framework with different visual feature configurations

Feature configuration	Identification accuracy (%)
Local + global features	69.7
Local features (4×4)	60.3
Global features	58.7

Table 6
Training time and identification accuracy comparison of our method with different numbers of hidden units in RBF classifier

Size of hidden unit	Training cost (s)	Query accuracy (%)
20	2340	69.7
17	2820	69.2
14	3980	69.7
10	4830	69.2
8	6920	69.1
6	9290	69.3
4	10 300	69.5

The our system was tested based on three different visual feature sets: local visual features, global visual features and local features plus global features. Table 5 summarizes the results of this study (of visual feature configurations). The main observation is that it is difficult to achieve superior identification accuracy with the system only considering local visual features or global visual features. For example, the system based on local features and based on global feature provide an identification accuracy of 60.8% and 58.7% individually. Comparing the system integrating both local and global visual features, there is a big room in terms of accuracy improvement.

7.2. Effects of parameters of neural network

The parameter setting enjoys great influence on data modeling effectiveness for neural network. It is very time consuming task to select the optimal parameter to maximize the performance. In fact, a wide variety of parameter values can be tuned in order to find an optimal choice for the network learning algorithm in the above experiments. However, performing a large number of random parameter tests is impossible in real application due to resource limitation or other reasons. Additionally, different applications may require different sets of parameters of the network. Since number of hidden layer is the important parameter for RBF neural network, we carry out the corresponding empirical study on this factor and basic methodology is to compare training cost and query accuracy of the system containing RBF network with different numbers of hidden units.

Table 6 summaries the training time of neural network and identification accuracy with various numbers of hidden units. It is not hard to see that the more hidden units the neural network contains, the training cost required to complete the learning process becomes more expensive. However, more complex neural network architecture does not guarantee the better identification accuracy. Since the network serves as a sub classifier, the number of hidden units is restricted to a practical limit in terms of computational cost. We find that 14 is optimal number for hidden unit based on this empirical study.

8. Conclusion and challenges ahead

With continued growth of image content available on the Internet, visual information processing has attracted significantly increasing attention from different research fields. Since art and culture heritage have always played an crucial role in our daily lives, classical painting categorization based on "artist" is being essentially important due to numerous potential applications in digital library and museum. In this article, we described a novel framework to facilitate effective artist identification process in large classic paint databases. The approach can effectively integrate both local and global visual information for identification process. To further enhance the system, we also design special encoding scheme called binary output codes with penalty (BOCP) to combine likelihood scores from various subclassifiers. The system has been fully implemented and tested with different data sets. As shown in experimental evaluation, our system significantly improves the effectiveness and efficiency over the state of art systems. At the same time, better robustness against visual noise can be achieved. The work related to our current approach continues along several directions. Automatic identification over media data is fundamental research problem and will enjoy greatest impact. In coming future, we plan to extend the current techniques to support identification problem for a wide range of media data. Also current evaluation procedure is very time consuming and tedious. Thus, developing analytical cost models that can accurately predict the query costs in terms of effectiveness, space and time complexity is another interesting research direction. Furthermore, until now there is no standard test collection for evaluating painter identification methods. We believe the test collection we create has great potential to be standard data set for unify evaluation process. Moreover, how to generating large scale testbeds is another open but promising research problem with great impact to the relative communities.

References

[1] C. Carson, S. Belongie, H. Greenspan, J. Malik, Blobworld: image segmentation using expectation–maximization and its application to image querying, IEEE Trans. Pattern Anal. Mach. Intell. 24 (8) (2002) 1026–1038.

[2] R. Datta, D. Joshi, J. Li, J.Z. Wang, Image retrieval: ideas, influences, and trends of the new age, ACM Comput. Surveys 40 (65) (2008).

[3] I. Cox, J. Ghosn, L. Miller, S.M. Omohundro, P. Yianilos, Pichunter: Bayesian relevance feedback for image retrieval, in: Proceedings of IEEE International Conference on Pattern Recognition, 1996, pp. 361–369.

[4] I. Cox, L. Miller, P. Minka, V. Papathomas, P. Yianilos, The Bayesian image retrieval system, pichunter: theory, implementation and psychophysical experiments, IEEE Trans. Image Process. 9 (1) (2000) 20–37.

[5] J. Li, N. Allinsion, D. Tao, X. Li, Multi-training support vector machine for image retrieval, IEEE Trans. Image Process. 15 (11) (2006).

[6] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, P. Yanker, Query by image and video content: the QBIC system, IEEE Comput. 28 (9) (1995) 23–32.

[7] A. Natsev, R. Rastogi, K. Shim, Walrus: a similarity retrieval algorithm for image databases, IEEE Trans. Know. Data Eng. 16 (3) (2004) 301–316.

[8] D. Tao, X. Tang, X. Li, Which components are important for interactive image searching?, IEEE Trans. Circuits Systems Video Technol. (TCSVT) 17 (2007).

[9] A.S. Harris, Seventeenth-Century Art and Architecture, Laurence King, 2005.

[10] G. Bazin, Baroque and Rococo (World of Art), Thames and Hudson, 1998.

[11] G. Davies, H. Eills, E. Shepherd, Perceiving and Remembering Faces, Academic, New York, 1981.

[12] R. Arnheim, Art and Visual Perception: A Psychology of the Creative Eye, University of California Press, Berkeley, CA, 2004.

[13] J. House, Monet in the 20th Century, Yale University Press, New Haven, CT, 1998.

[14] T. Melzer, P. Kammerer, E. Zolda, Stroke detection of brush strokes in portrait miniatures using a semi-parametric and a model based approach, in: Proceedings of IEEE ICPR, 1998.

[15] R. Sablatnig, P. Kammerer, E. Zolda, Hierarchical classification of paintings using face- and brush stroke models, in: Proceedings of IEEE ICPR, 1998.

[16] C. Colombo, A.D. Bimbo, P. Pala, Semantic in visual information retrieval, IEEE Multimedia 6 (3) (1999) 38–53.

[17] J. Li, J.Z. Wang, Studying digital imagery of ancient paintings by mixtures of stochastic models, IEEE Trans. Image Process. 13 (3) (2004) 340–353.

[18] S. Jiang, Q. Huang, Q. Ye, W. Gao, An effective method to detect and categorize digitized traditional Chinese paintings, Pattern Recognition Lett. 27 (7) (2006) 734–746.

[19] I. Berezhnoy, E. Postma, J. Herik, Computerized visual analysis of paintings, in: Proceedings of International Conference on Association for History and Computing, 2005.

[20] D.A. Forsyth, J. Ponce, Computer Vision: A Modern Approach, Prentice-Hall, Englewood Cliffs, NJ, 2003.

[21] R.C. Gonzalez, R.E. Woods, Digital Image Processing, Prentice-Hall, Englewood Cliffs, NJ, 2002.

[22] A. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, Content-based image retrieval at the end of early year, IEEE Trans. Pattern Anal. Mach. Intell. 22 (12) (2000) 1349–1380.

[23] T.S. Lee, Image representation using 2d gabor wavelets, IEEE Trans. Pattern Anal. Mach. Intell. 18 (10) (1996) 959–971.

[24] R. Myers, Classical and Modern Regression with Application, Durbury Press, 1989.

[25] C.M. Bishop, Neural Network for Pattern Recognition, Oxford University Press, Oxford, 2000.

[26] T.G. Dietterich, G. Bakiri, Solving multiclass learning problems via error-correcting output codes, J. Artif. Intell. Res. 2 (1995) 263–286.

[27] R.W. Hamming, Error detecting and error correcting codes, Bell System Techn. J. 26 (2) (1950) 147–160.

[28] D. Montgomery, E.A. Peck, G.G. Vining, Introduction to Linear Regression Analysis, Wiley, New York, 2001.

About the Author—JIALIE SHEN is currently an Assistant Professor at Singapore Management University, Singapore. His research interests can be summarized as developing effective and efficient data analysis and retrieval techniques for novel data intensive applications. Particularly, he is currently interested in various techniques of multimedia data mining, multimedia information retrieval and database systems. His research results have been published in ACM SIGIR, ACM SIGMOD, ACM Multimedia, ICDE, CIKM and IEEE Transactions on Multimedia. He has served as a Reviewer for a number of major journals and conferences such as ICDE, IEEE TMM, IEEE TKDE and ACM TOIS. He is also a member of ACM SIGMOD and SIGIR.