INTERNATIONAL JOURNAL OF ELECTRONICS; MECHANICAL and MECHATRONICS ENGINEERING Vol.2 Num.2 pp.(215-224)

AUTHENTICATION OF OTTOMAN ART CALLIGRAPHERS

Osman N. Ucan[1] Mustafa Istanbullu[1] Niyazi Kilic[2] Ahmet Kala[3]

[1]Istanbul Aydin University, Engineering&Architecture Faculty, Electrical and Electronics Eng. Dept. Florya Istanbul, Turkey
uosman@aydin.edu.tr mustafaistanbullu@aydin.edu.tr
[2]Istanbul University Engineering Faculty, Electrical and Electronics Eng. Dept. 34320, Avcilar, Istanbul, Turkey
niyazik@istanbul.edu.tr
[3]Istanbul University Economy Faculty, Beyazit, Istanbul, Turkey
akala@istanbul.edu.tr

Abstract: *A content-based retrieval system can provide efficient access to the documents from the Ottoman Empire archives, which contain more than 100 million handwritten files. Experts want to archive the historical documents to be stored in digital image forms because the documents include not only text but also drawings, portraits, miniatures, signs, ink smears, etc., which might have an associated historical value. This work will help us to identify the Ottoman period calligraphies; accordingly determine its value and chronology. Here, we describe and apply a computational technique on high-resolution scanned form of the original works, specifically Ottoman art calligraphies for authentication. We show preliminary results from 9 well known Ottoman art calligraphers at various times. Since paintings have high resolution, complexity and capacity, it is not possible to process them in one block. In this paper, Wavelet Transform (WT) and Support Vector Machine (SVM) are employed in cascade form to obtain an objective approach in authentication of calligraphies, and we have succeed up to 100% percentage for classify the art works using these techniques.*

Keywords: *Art authentication, Ottoman art calligraphers, Wavelet transform, Support vector machine*

1. INTRODUCTION

Authentication of paintings, calligraphies can be extremely challenging. Typically, art experts reach decisions after thorough consideration of many different types of evidence. Artist's lifetime and documents and tracing the art's history of ownership provide clues. In addition to the reliance on the human actor, quantitative methods can be brought to bear. Available techniques of examination are given below: Carbon-14 dating is used to measure the age of an object up to 10,000 years old [1]. White lead dating is used to pinpoint the age of an object up to 1,600 years old [2]. Conventional X-ray can be used to detect earlier work present under the surface of a painting. Sometimes artists will legitimately re-use their own canvasses, but if the painting on top is supposed to be from the 17th century, but the one underneath shows people in 19th century properties such as dress models, the scientist will assume the top painting is not authentic. Also x-rays can be used to view inside an object to determine if the object has been altered or repaired. X-ray diffraction (the object bends X-rays) is used to analyze the components that make up the paint an artist used, and to detect pentimenti. X-ray fluorescence (bathing the object with radiation causes it to emit X-rays) can reveal if the metals in a metal sculpture or if the composition of pigments is too pure, or newer than their supposed age [3]. Quantitative techniques could potentially be helpful in identifying images, copies or forgeries thus investigators may rely on other methods such as computerized authentication. In literature, regarding to this case, there are some works based on image processing. [4, 5] In view of all these things, there is not so much working about Ottoman art calligraphies for authentication. In this case we think that this work will be an important point for authentication of Ottoman art calligraphies. Wavelet Transform is a mathematical tool, which analyses and decomposes one and two dimensional signals. Wavelet decomposition divides signals into sub-bands. These sub-bands are evaluated to determine textures, assigning a frequency to each sub-band and related coefficients of sub-bands can be utilized as feature extraction block and followed by a classifier. Then, these data are used as input into Support Vector Machine for testing [6].

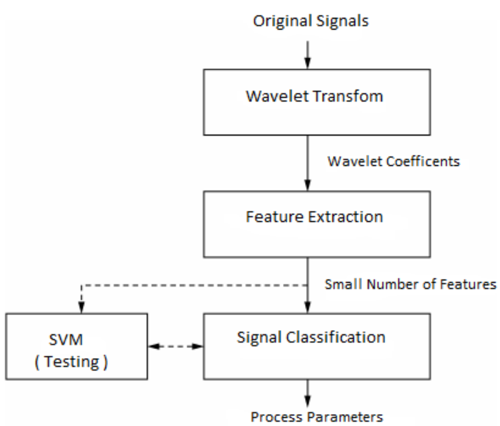

Figure 1 Computation Scheme

Here some important ottoman art calligraphers and their works have been chosen. Abbaskamil, Hafiz Rasihibrali, Hasan Rıza, İsmet el-Mevlevi, Mahmud Celaleddin, Sabri, Seyyid Derviş Halil, Seyyid Muhammed Bahir and Veliyyuddin are all milestones for art calligraphies of Ottoman Empire.

All this chosen ones are well known in their life period and have a lot of distinctive calligraphies. We have chosen their 50 art calligraphies for this work. These are some examples of the chosen ones:

Figure 1 Some calligraphy examples of Ottoman Empire

These art calligraphies and their digital photographs are separated by two consecutive approaches; wavelet and Support Vector Machine (SVM) in evaluation of high capacity 2D painting images. In high resolution images, one step direct image processing is almost impractical, because of time consuming long iterations. The first approach is based on compression with minimum loss via feature extraction. The second is classification of original art calligraphies and its fakes by SVM, where extracted features are taken as input. SVM developed by Vapnik [7, 8] have been used in a range of problems including pattern recognition, bioinformatics and text categorization [9-10]. SVM provides a novel approach to the two and multi-class classification problem.

2. METHODS

In this work, calligraphies some are given in Figure 2, segmented properly and samples augmented to 50 pieces. Then these pieces applied wavelet transform for feature extraction. In this step, wavelet coefficients of 50 images are calculated. At the first stage, 8 wavelet transform outputs; approximation, vertical, horizontal, diagonal of Level-1 DAUB4 and HAAR type wavelet decompositions and maximum, minimum, mean, variance, standard deviation values of raw image are obtained. In the second stage of cascade structure, these 13 parameters are chosen as input of SVM for classification. As a result, authentication accuracy of this cascade block, has reached up to 100%.

2.1. Wavelet Based Feature Extractıon

Wavelets decompose data into different frequency sub-bands components and then study each component with a resolution matched to its scale. Wavelets have come out as powerful new mathematical tools for analysis of complex datasets. In classical approach, Fourier transform provides representation of an image based only on its frequency contents. Hence this representation is not spatially localized while wavelet functions are localized in space. While Fourier transform groups a signal into a spectrum of frequencies whereas the wavelet analysis decomposes into a hierarchy of scales ranging from the coarset scale. Hence wavelet transform which provides representation of an image at various resolutions is a better tool for feature extraction from images [11, 12]. The wavelet transform is a useful mathematical tool that currently has received a great attention in different applications like compression and feature extraction. Feature extraction is defined that extraction some important features from the image and obtaining feature vector [13]. Here, wavelet transform is used as a feature extractor.

In numerical analysis and functional analysis, a discrete wavelet transform (DWT) is any wavelet transform for which the wavelets are discretely sampled. Therefore, since our data is discrete, we use Discrete Wavelet Transform. DWT employs a discrete set of the wavelet scales and translation obeying some defined rules as an implementation of the wavelet transform. As with other wavelet transforms, a key advantage it captures both frequency and location information. Here, both of scale and translation parameters are discrete. Thus, DWT can be represented in Eq.(1),

$$W[m,n] = \sum_{x} f[x]\psi_{m,n}[x] \tag{1}$$

where, discretized scale and translation parameters are given by, $a = 2^j$ ve $b = k2^j$ ($k, j \in Z$). Then, wavelet basis function is writen in Eq.(2),

$$\psi_{j,k}[x] = 2^{-j/2}\psi(2^j x - k) \tag{2}$$

One dimensional transforms are easily extended to two dimensional functions like images. In this case, the DWT is applied to each dimension separately. This yield a multi resolution decomposition of the image into four sub-bands called the approximation (low frequency component) and details (high frequency component). The approximation (A) indicates a low resolution of the original image. The detail coefficients are horizontal (H), vertical (V), and diagonal (D).

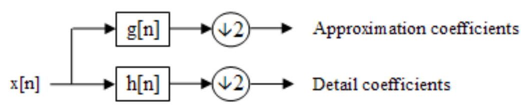

Figure2 DWT Transform

2.2. Support Vector Machine Based Classfication

Support vector machines are a set of relates supervised learning methods that analysis data and recognize patterns, used for classification. The original SVM algorithm was invented by Vladimir Vapnik and current standard incarnation (soft margin) was purposed by Corinna Cortes and Vladimir Vapnik. The standard SVM takes a set of input data and predicts, for each given input, which of two possible classes the input is a member of, which makes the SVM a non-probabilistic binary linear classifier. It is formulized under the concept of structural risk minimization rule unlike neural network based classifier [10]. At the beginning, SVM was actually designed for binary classification in order to construct an optimal hyperplane so that the margin of separation between the negative and positive data set will be maximized. Let $\{x_i, y_i\}, i = 1,2,.....l, \quad y_i \in \{-1,1\}$ and $x_i \in IR^n$ be the training samples where the training vector is x_i and y_i is its corresponding labeled class. As a result SVM can be expressed as follows;

$$f(x) = sign\left(\sum_{i=1}^{l} y_i \alpha_i K(x, x_i) + b\right) \tag{3}$$

Where,

$$\operatorname{sgn}(u) = \begin{cases} 1 & for \;\; u > 0 \\ -1 & for \;\; u < 0 \end{cases}$$

where l is the number of learning patterns, y_i is the target value of learning pattern x_i , b is a bias, and $K(x, x_i)$ is a kernel function that high-dimensional feature space:

$$K(x, x_i) = \phi(x).\phi(x_i) \tag{4}$$

The polynomial kernel which is shown in equation (5) and the Gaussian radial basis function (RBF) kernel which is shown in equation (6) are frequently used kernel functions.

$$K(x, x_i) = (x.x_i + 1)^p \tag{5}$$

$$K(x,x_i) = \exp[\gamma |x - x_i|] \tag{6}$$

In polynomial kernel, p is the degree of polynomial kernel; if p is equal to 1, kernel is called linear and if it is equal to 2, then kernel is called quadratic kernel.

2.2.1. Training of Support Vector Machine

SVM finds the hyperplane that causes maximizes the separating margin between two positive and negative classes. Mathematically, this hyperplane can be found by minimizing the following cost function,

$$P(w) = \frac{1}{2}\|w\|^2 \tag{7}$$

Subject to $y_i.f(x_i) \geq +1 - \varepsilon_i \quad \varepsilon_i \geq 0, \quad i = 1,2,....,k$ the aim is to maximize the separating margin subject to constraints. This problem is transformed to dual form for solving which follows:

$$L_D = \sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i x_j \tag{8}$$

Subject to $0 \leq \alpha \leq C$ and $\sum_i \alpha_i y_i = 0$.

Where, C is regularization parameter that controls the tolerance to classification errors in training. The training vector x_i whose corresponding α_i is nonzero is called support vector.

3. AUTHENTICATION OF ART CALLIGRAPHIES

Wavelet coefficients of 50 images are calculated. After decomposition with Discrete Wavelet Transform, we have four coefficients for each image; approximation, vertical, horizontal and diagonal. Level-1 DAUB4 and HAAR type wavelet decompositions are preferred. The first level decomposition vector size is too large to be given as an input to a classifier. Since high dimensional of feature vectors increased computational complexity and hence, in order to reduce to dimensionality of the extracted feature vectors, statistics over the wavelet coefficients are used. The statistical features were also chosen as; maximum, minimum, mean, variance, standard deviation values of {approximation, vertical, horizontal and diagonal} sub bands. The computed statistical features of 13 discrete feature coefficients are used as the inputs of the network of SVM. In SVM training, linear, quadratic and RBF kernel is used as in Tables 1-5. General review is shown in table 5. The best error level is achieved when RBF kernel parameters.

4. CONCLUSION

In this study, we have classified calligraphies of 9 Ottoman calligraphers and its images with the help of wavelet transform and SVM, which are used for feature extraction and supervised machine learning approaches. We have

preferred wavelet type is Level-1 Daub4 and Haar after experimental study. For compression and unification, only some statistics of first level wavelet coefficients are used as an input of SVM. These are; {maximum, minimum, mean, variance and standard deviation values of the {approximation, vertical, horizontal and diagonal} sub bands.
Our experimentation showed that classification accuracy has reached up to 100% by Daub4 type Wavelet and cascaded by RBF Kernel Type SVM. This statistic suggests that our combined approach may be used to facilitate separation from original painting to its fake.

5. REFERENCES

[1] http://amath.colorado.edu/courses/2460/2006Spr/Lab1/Forgeries.pdf, Detecting Art Forgeries

[2] http://www.mystudios.com/artists-mind/volume-04-2002.html

[3] LYU S., D. ROCKMORE, H. FARID (2004) A digital technique for art authentication, Proceeding of the National Academy of Sciences, 101 (49):17006-17010.

[4] ŞAYKOL E., SINOP A. K. , GÜDÜKBAY U., ULUSOY Ö., ÇETIN E., Content-Based Retrieval of Historical Ottoman Documents Stored as Textual Images, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 13, NO. 3, MARCH 2004, p: 314-325

[5] GUNSEL B., S. SARIEL, O. ICOGLU (2005) Content-based access to art paintings, Proc. of IEEE ICIP, Genova, Italy, September 2005.

[6] KIRAN K. SIMHADRI, S. S. IYENGAR, Ronald J. HOLYER, Matthew LYBANON, John M. ZACHARY, (1998) Wavelet-Based Feature Extraction from Oceanographic Images, Ieee Transactions on Geoscience and Remote Sensing, Vol. 36, No. 3, May 1998, p: 767-778

[7] VAPNIK V (1999) An overview of statistical learning theory, IEEE Transaction on Neural Network, 10:989-999.

[8] POLAT K., S.GUNEŞ (2007) Breast cancer diagnosis using least square support vector machine, Digital Signal Processing, 17:694-701.

[9] BURGES C (1998) A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery, 2:1-43.

[10] DAUBECHIES I (1992) Ten lectures on wavelets, CBMS Conference Lecture Notes 61. SIAM. Philadelphia.

[11] GONZALES, R.C., R.E WOODS (1993) Digital Image Processing, Addison Wesley, Reading, MA.

[12] DEVIJVER, P.A., J., KITTLER (1982) Pattern Recognition: A Statistical Approach, Prentice-Hall, London.

[13] ŞENGÜR A. , TÜRKOĞLU I. (2004) Değişmez Momentlerle Türkçe Karakter Tanıma, Doğu Anadolu Bölgesi Araştırmaları.

6. LIST OF TABLES

Model 1: In 9 categories, obtaning results using only Appr coefficents and db4 wavelet for 50 data: (SVM, 5 fold, c=10, RBF gama:0,1)

Table 1 a) General performance b) Performance of classification due to calligraphers c) Confusion matrix

Correctly Classified Instances	44 (88%)
Incorrectly Classified Instances	6 (12%)
Kappa Statistic	0.8644
Mean Absolute Error	0.1747
Root Mean Squared Error	0.2847
Relative Absolute Error	88.59%
Root Relative Squared Error	90.60%
Total Number of Instances	50

a)

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Abbaskamil	1	0	1	1	1	1
Hafızrasihibra	0.667	0.045	0.667	0.667	0.667	0.82
Hasanrıza	1	0	1	1	1	1
İsmetelmevlevi	0.8	0	1	0.8	0.889	0.996
Mahmudcelaleddin	1	0	1	1	1	1
Sabri	1	0.022	0.833	1	0.909	0.989
Seyyiddervişhalil	0.875	0.024	0.875	0.875	0.875	0.985
Seyyidmuhammedbahir	0.833	0	1	0.833	0.909	0.875
Veliyyuddin	0.8	0.044	0.667	0.8	0.727	0.964
Weighted Avg.	0.88	0.016	0.89	0.88	0.882	0.956

b)

CLASSIFIED AS	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**
a =Abbaskamil	5	0	0	0	0	0	0	0	0
b = Hafızrasihibrali	0	4	0	0	0	0	1	0	1
c = Hasanrıza	0	0	6	0		0	0	0	0
d = İsmetelmevlevi	0	0	0	4	0	1	0	0	0
e = Mahmudcelaleddin	0	0	0	0	4	0	0	0	0
f = Sabri	0	0	0	0	0	5	0	0	0
g = Seyyiddervişhalil	0	1	0	0	0	0	7	0	0
h = Seyyidmuhammedbahir	0	0	0	0	0	0	0	5	1
i = Veliyyuddin	0	1	0	0	0	0	0	0	4

c)

Model 2: In 9 categories, obtaning results using only Appr coefficents and Haar wavelet for 50 data (SVM 5 fold, RBF gama:0,1)

Table 2 a) General performance b) Performance of classification due to calligraphers c) Confusion matrix

Correctly Classified Instances	43 (86%)
Incorrectly Classified Instances	7 (14%)
Kappa Statistic	0.8418
Mean Absolute Error	0.1749
Root Mean Squared Error	0.2851
Relative Absolute Error	88.72%
Root Relative Squared Error	90.73%
Total Number of Instances	50

a)

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Abbaskamil	0.8	0	1	0.8	0.889	0.989
Hafızrasihibrali	0.667	0.045	0.667	0.667	0.667	0.82
Hasanrıza	1	0	1	1	1	1
İsmetelmevlevi	0.8	0	1	0.8	0.889	0.993
Mahmudcelaleddin	1	0	1	1	1	1
Sabri	1	0.022	0.833	1	0.909	0.989
Seyyiddervişhalil	0.875	0.024	0.875	0.875	0.875	0.985
Seyyidmuhammedbahir	0.833	0.023	0.833	0.833	0.833	0.85
Veliyyuddin	0.8	0.044	0.667	0.8	0.727	0.964
Weighted Avg.	0.86	0.019	0.87	0.86	0.861	0.952

b)

CLASSIFIED AS	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**
a =Abbaskamil	4	0	0	0	0	0	0	1	0
b = Hafızrasihibrali	0	4	0	0	0	0	1	0	1
c = Hasanrıza	0	0	6	0	0	0	0	0	0
d = İsmetelmevlevi	0	0	0	4	0	1	0	0	0
e = Mahmudcelaleddin	0	0	0	0	4	0	0	0	0
f = Sabri	0	0	0	0	0	5	0	0	0
g = Seyyiddervişhalil	0	1	0	0	0	0	7	0	0
h = Seyyidmuhammedbahir	0	0	0	0	0	0	0	5	1
i = Veliyyuddin	0	1	0	0	0	0	0	0	4

c)

Model 3: In 9 categories, obtaning results, taking statistics of appr and detail coefficents and using db4 wavelet for 50 data (SVM, RBF, gama:0,1)

Table 3 a) General performance b) Performance of classification due to calligraphers c) Confusion matrix

Correctly Classified Instances	50(100%)
Incorrectly Classified Instances	0 (0%)
Kappa Statistic	1
Mean Absolute Error	0.1728
Root Mean Squared Error	0.2814
Relative Absolute Error	87.65 %
Root Relative Squared Error	89.55 %
Total Number of Instances	50

a)

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Abbaskamil	1	0	1	1	1	1
Hafızrasihibrali	1	0	1	1	1	1
Hasanrıza	1	0	1	1	1	1
İsmetelmevlevi	1	0	1	1	1	1
Mahmudcelaleddin	1	0		1	1	1
Sabri	1	0	1	1	1	1
Seyyiddervişhalil	1	0	1	1	1	1
Seyyidmuhammedbahir	1	0	1	1	1	1
Veliyyuddin	1	0	1	1	1	1
Weighted Avg.	1	0	1	1	1	1

b)

CLASSIFIED AS	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**
a =Abbaskamil	5	0	0	0	0	0	0	0	0
b = Hafızrasihibrali	0	6	0	0	0	0	0	0	0
c = Hasanrıza	0	0	6	0	0	0	0	0	0
d = İsmetelmevlevi	0	0	0	5	0	0	0	0	0
e = Mahmudcelaleddin	0	0	0	0	4	0	0	0	0
f = Sabri	0	0	0	0	0	5	0	0	0
g = Seyyiddervişhalil	0	0	0	0	0	0	8	0	0
h =Seyyidmuhammedbahir	0	0	0	0	0	0	0	6	0
i = Veliyyuddin	0	0	0	0	0	0	0	0	5

c)

Model 4: In 9 categories, obtaning results, taking statistics of appr and detail coefficents and using Haar wavelet for 50 data:

Table 4 a) General performance b) Performance of classification due to calligraphers c) Confusion matrix

Correctly Classified Instances	50(100%)
Incorrectly Classified Instances	0 (0%)
Kappa Statistic	1
Mean Absolute Error	0.1728
Root Mean Squared Error	0.2814
Relative Absolute Error	87.65 %
Root Relative Squared Error	89.56 %
Total Number of Instances	50

a)

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Abbaskamil	1	0	1	1	1	1
Hafızrasihibrali	1	0	1	1	1	1
Hasanrıza	1	0	1	1	1	1
İsmetelmevlevi	1	0	1	1	1	1
Mahmudcelaleddin	1	0	1	1	1	1
Sabri	1	0	1	1	1	1
Seyyiddervişhalil	1	0	1	1	1	1
Seyyidmuhamme ahir	1	0	1	1	1	1
Veliyyuddin	1	0	1	1	1	1
Weighted Avg.	1	0	1	1	1	1

b)

CLASSIFIED AS	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**
a =Abbaskamil	5	0	0	0	0	0	0	0	0
b = Hafızrasihibrali	0	6	0	0	0	0	0	0	0
c = Hasanrıza	0	0	6	0	0	0	0	0	0
d = İsmetelmevlevi	0	0	0	5	0	0	0	0	0
e = ahmudcelaleddin	0	0	0	0	4	0	0	0	0
f = Sabri	0	0	0	0	0	5	0	0	0
g = Seyyiddervişhalil	0	0	0	0	0	0	8	0	0
h =Seyyidmuhammedbahir	0	0	0	0	0	0	0	6	0
i = Veliyyuddin	0	0	0	0	0	0	0	0	5

c)

Model-1, only wavelet appr coefficients and db4 wavelet; ***Model-2,** only wavelet appr coefficients and Haar wavelet; ***Model-3,** some statistics (min, max, mean, std, moment3) of wavelet coefficients (appr and detail) and db4 wavelet; ***Model-4,** some statistics (min, max, mean, std, moment3) of wavelet coefficients (appr and detail) and Haar wavelet

Table 5 General review of all models

Classificatory	Model	Accurate Of Classification (%)
SVM	Model-1*	88
9 category	Model-2*	86
50 data	Model-3*	100
	Model-4*	100