Arabic Character Recognition Using Regression Curves with the Expectation Maximization Algorithm

Abdullah A. AlShaher

Abstract—In this paper, we demonstrate how regression curves can be used to recognize 2D non-rigid handwritten shapes. Each shape is represented by a set of non-overlapping uniformly distributed landmarks. The underlying models utilize 2nd order of polynomials to model shapes within a training set. To estimate the regression models, we need to extract the required coefficients which describe the variations for a set of shape class. Hence, a least square method is used to estimate such modes. We then proceed by training these coefficients using the apparatus Expectation Maximization algorithm. Recognition is carried out by finding the least error landmarks displacement with respect to the model curves. Handwritten isolated Arabic characters are used to evaluate our approach.

Keywords—Shape recognition, Arabic handwritten characters, regression curves, expectation maximization algorithm.

I. INTRODUCTION

SHAPE recognition has been the focus of many researchers since seven past decades [1] and attracted many communities in the field of pattern recognition [2], artificial intelligence[3], signal processing [4], image analysis [5], and computer vision [6]. The difficulties arise when the shape under study exhibits high degree in shape variation: as in handwritten characters [7], digits [8], face detection [9], and gesture authentication [10]. For a single data, shape variation is limited and cannot be captured ultimately due to the fact that single data does not provide sufficient information and knowledge about the data; therefore, multiple existence of data provides better understanding of shape analysis and manifested by mixture models [11]. Because of the existence of multivariate data under study, there is always the requirement to estimate the parameters that describe the data that is encapsulated within a mixture of shapes.

The literature demonstrates many statistical and structural approaches with various algorithms to model shape variations using supervised and unsupervised learning [12] algorithms. In precise, the powerful Expectation Maximization algorithm of Dempster [13] comes that has been used widely for such cases. The EM algorithm revolves around two step procedure. The expectation E step revolves around estimating the parameters of a log-likelihood function and pass it to the Maximization M step. In a maximization (M) step, the algorithm computes parameters maximizing the expected log-likelihood found on the E step. The process is iterative one until all parameters comes to unchanged. For instance, Jojic and Frey [14] have used the EM algorithm to fit mixture models to the appearance manifolds for faces. Bishop and Winn [15] have used a mixture of principal components analyzers to learn and synthesize variations in facial appearance. Vasconcelos and Lippman [16] have used the EM algorithm to learn queries for content-based image retrieval. Finally, several authors have used the EM algorithm to track multiple moving objects [17]. Revov et al. [18] have developed a generative model which can be used for handwritten character recognition. Their method employs the EM algorithm to model the distribution of sample points.

Curves are used widely by research in the computer vision society [1]-[5]. Curvatures are mainly used to distinguish different shapes such as characters [6], digits, faces [2], and topographic maps [3]. Curve fitting [18], [19] is the process of constructing a 2nd order or higher mathematical function that has the best fit to a series of landmark points. A related topic is regression analysis that stresses on probabilistic conclusion on how uncertainty is present when fitting a curve to a set of data landmarks with marginal errors. Regression curves are applied for data visualization [12], [13] to capture the values of a function with missing data [14] and to gain the relationship of multiple variables.

In this paper, we demonstrate how curves are used to recognize 2D handwritten shapes by fitting 2nd order of polynomial quadratic function to a set of landmarks points presented in a shape. We then train such curves to capture the optimal characteristics of the shapes in the training sets of curves. Handwritten Arabic characters are used and tested in this investigation.

II. REGRESSION CURVES

We would like to extract the best fit modes that describe the shapes under study, hence, a multiple image shapes are required and is explained by a training sets of class shape Ω and the complete sets of shape classes denoted by Ω. Let us assume that each training set is represented by the following 2D training patterns as a long vector

\[
X^Ω = (x^0_1, y^0_1, x^0_2, y^0_2, \ldots, x^0_{k}, y^0_{k}), (x^{a_1}_1, y^{a_1}_1, x^{a_1}_2, y^{a_1}_2, \ldots, x^{a_1}_{k}, y^{a_1}_{k}, x^{a_2}_1, y^{a_2}_1, x^{a_2}_2, y^{a_2}_2, \ldots, x^{a_2}_{k}, y^{a_2}_{k}, \ldots)
\]

Our model here is a polynomial of higher order. In this case, we choose 2nd order of quadratic curves. Consider the following generic form of a polynomial of order j

Abdullah A. AlShaher is with the Department of Computers and Information Systems, College of Business Studies, The Public Authority for Applied Education and Training, Ardiyah, The State of Kuwait (e-mail: aa.alshaher@paaet.edu.kw).
\( f(x_{\omega}) = a_0 + a_1x_{\omega}^{\alpha_1} + a_2x_{\omega}^{\alpha_2} + a_3(x_{\omega}^{\alpha_3})^2 + \ldots + a_T(x_{\omega}^{\alpha_T})^T = a_0 + \sum_{t=1}^{T}ax_{\omega}^{\alpha_t} \) (2)

The nonlinear regression above requires the estimation of the coefficients that best fit the sample shape landmarks, we approach the least square error between data \( y \) and \( f(x) \) in

\[
\text{err} = \sum_{i=1}^{T}(y_{\omega}^{\alpha_t} - (a_0 + \sum_{k=1}^{K}a_kx_k^{\beta}))^2
\]

where \( T \) is the number of pattern set, \( k \) is the current data landmarks point being summed, \( j \) is the order of polynomial equation. Rewriting (14) in a more readable format

\[
\text{err} = \sum_{k=1}^{K}(y_{\omega}^{\alpha_t} - (a_0 + \sum_{k=1}^{K}a_kx_k^{\beta}))^2
\] (5)

Finding the best fit curve is equivalent to minimizing the squared distance between the curve and landmark points. The aim here is to find the coefficients, hence, solving the equations of taking the partial derivative with respect each coefficient \( a_0, a_1 \); for \( k = 1, \ldots, j \) and set each to zero in

\[
\frac{\partial \text{err}}{\partial a_0} = \sum_{k=1}^{K}(y_{\omega}^{\alpha_t} - (a_0 + \sum_{k=1}^{K}a_kx_k^{\beta})) = 0
\] (6)

\[
\frac{\partial \text{err}}{\partial a_1} = \sum_{k=1}^{K}(y_{\omega}^{\alpha_t} - (a_0 + \sum_{k=1}^{K}a_kx_k^{\beta}))x_k^{\beta} = 0
\] (7)

\[
\frac{\partial \text{err}}{\partial a_2} = \sum_{k=1}^{K}(y_{\omega}^{\alpha_t} - (a_0 + \sum_{k=1}^{K}a_kx_k^{\beta}))x_k^{2\beta} = 0
\] (8)

Rewriting upper equations in the form of matrix and applying linear algebra matrix differentiation, we get

\[
\begin{bmatrix}
T \\
\sum_{k=1}^{K}x_k^{\alpha_1} & \sum_{k=1}^{K}x_k^{\alpha_2} & \ldots & \sum_{k=1}^{K}x_k^{\alpha_T} \\
\sum_{k=1}^{K}x_k^{\alpha_2} & \sum_{k=1}^{K}x_k^{\alpha_3} & \ldots & \sum_{k=1}^{K}x_k^{\alpha_T} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{k=1}^{K}x_k^{\alpha_T} & \sum_{k=1}^{K}x_k^{\alpha_{T+1}} & \ldots & \sum_{k=1}^{K}x_k^{\alpha_{T+T}}
\end{bmatrix}
\begin{bmatrix}
a_0 \\
a_1 \\
a_2 \\
\vdots \\
a_T
\end{bmatrix}
= \begin{bmatrix}
\sum_{k=1}^{K}y_k^{\alpha_1} \\
\sum_{k=1}^{K}y_k^{\alpha_2} \\
\vdots \\
\sum_{k=1}^{K}y_k^{\alpha_T}
\end{bmatrix}
\] (9)

Choosing Gaussian elimination procedure to rewrite the upper equation in more solvable in

\[
Ax = B
\] (10)

where

\[
A = \begin{bmatrix}
T \\
\sum_{k=1}^{K}x_k^{\alpha_1} \\
\sum_{k=1}^{K}x_k^{\alpha_2} \\
\vdots \\
\sum_{k=1}^{K}x_k^{\alpha_T}
\end{bmatrix} \quad X = \begin{bmatrix}
a_0 \\
a_1 \\
\vdots \\
a_T
\end{bmatrix}, \quad B = \begin{bmatrix}
\sum_{k=1}^{K}y_k^{\alpha_1} \\
\sum_{k=1}^{K}y_k^{\alpha_2} \\
\vdots \\
\sum_{k=1}^{K}y_k^{\alpha_T}
\end{bmatrix}
\] (11)

solving for \( X \) to find the coefficients \( A, B \) in

\[
X = A^{-1} \ast B
\] (12)

The outcome would be the coefficients \( a_0, a_1, a_2 \). We follow the similar procedure to find the remaining coefficients of the landmarks points.

### III. LEARNING REGRESSION CURVES

It has been known that when using learning algorithms to train models of such a case, the outcome is trained models with superior performance than those of untrained models bishop [19]. In this stage, we are concerned by capturing the optimal curve coefficients which describe the pattern variations under testing; hence, training is required by fitting the Gaussian mixtures of curve coefficient models to set of shape curve patterns. The previous approaches regard producing variations in shapes that of a linear fashion. To produce more complex shape variations, we have to proceed by employing non-linear deformation of a set of curve coefficients. Unsupervised learning is used encapsulated within a framework of the apparatus Expectation Maximization EM algorithm. The idea is borrowed from Cootes [20] of constructing point distribution models; however, the algorithm is transformed to learn regression curves coefficients \( a_t \) similar to that approach of Alsaher [21]. Suppose that a set of curve coefficients \( a_t \) for a set of training patterns is \( t = (1 \ldots T) \) where \( T \) is the complete set of training curves is represented in a long vector of coefficients :

\[
a_t = (a_{t_1}, a_{t_2}, a_{t_3}, a_{t_4}, a_{t_5}, a_{t_6}, \ldots, a_{t_n})\]

The mean vector of coefficient patterns is represented by

\[
\mu = \frac{1}{T} \sum_{t=1}^{T} a_t
\] (14)

The covariance matrix is then constructed by

\[
\Sigma = \frac{1}{T} \sum_{t=1}^{T} (a_t - \mu)(a_t - \mu)^T
\] (15)

The following approach is based on fitting a Gaussian mixture models to the set of training examples of curve coefficients. We further assume that training patterns are independent from one each other; thus, they are neither flagged nor labelled to any curve class. Each curve class \( \omega \) belongs to the set of curve classes \( \Omega \) has its own mean \( \mu \) and covariance matrix \( \Sigma \). With these ingredients, we establish the likelihood function for the set of the curve patterns in

\[
p(a_t) = \prod_{t=1}^{T} \sum_{\omega=1}^{\Omega} p(a_t | \mu_{\omega}, \Sigma_{\omega})
\] (16)

where the term \( p(a_t | \mu_{\omega}, \Sigma_{\omega}) \) is the probability for drawing curve pattern \( a_t \) from the curve-class \( \omega \). Associating the above likelihood function with the Expectation Maximization algorithm, the likelihood function can be written to be iterative process of two steps. The process revolves around estimating the expected log-likelihood function iteratively in
where the quantity and are the estimated mean curve vector and variance covariance matrix both at iteration (n) of the algorithm. The quantity is the a posteriori probability that the training pattern curve belong to the curve-class ω at iteration n of the algorithm. The term is the probability of distribution of curve-pattern αt belonging to curve-class ω at iteration (n + 1) of the algorithm; thus, the probability density to associate curve-patterns at iteration n along with the mixing proportion parameters for curve-class ω.

In the E, or Expectation step of the algorithm, the a posteriori curve-class probability is updated by applying the Bayes factorization rule to the curve-class distribution density at iteration n + 1. The quantity is the probability of distribution of curve-classes Ω at iteration n along with the mixing proportion parameters for curve-class ω.

where the revised curve-class ω mixing proportions are computed and ready to be used for recognition.

Both E, and M steps are iteratively converged, the outcome of the learning stage is a set of curve-class ω parameters such as , hence the complete set of all curve-class Ω are computed and ready to be used for recognition.

IV. RECOGNITION

In this stage, we focus of utilizing the parameters extracted from the learning phase for the purpose of shape recognition. Here, we assume that the testing shapes are classified based on the computing the new point position of testing data χ after projecting the sequence of curve-coefficients to the testing data in

Such testing patterns are classified according to the computing the highest probability using Bayes rule over the total curve-classes Ω in

V. EXPERIMENTS

We have evaluated our approach with sets of Arabic handwritten characters. Here, we have used 23 shape-classes for different writers, each with 80 training patterns. In total, we have tested the approach with 1840 handwritten Arabic character shape patterns for testing and 4600 patterns for testing phase. Figs. 1 and 2 show some training patterns used in this paper. Fig. 3 shows single shapes and their landmarks representation.

Fig. 4 demonstrates regression sample curve-classes as a result of the training stage. Fig. 5 demonstrate the curve-classes Ω convergence rate graph as a function per iteration no. is the training phase. The graphs show how associated distributed probabilities for the set of curve-classes Ω converged in a few iterations.
To take the investigation further, we demonstrate how well the approach behaves in the presence of noise. In figure 6, we show how recognition rate is achieved when point position displacement error is applied. Test shape coordinates are being moved away from their original position. The figure shows the recognition rate fails to register shapes to their correct class in a few iterations and it fails completely when coordinates are moved away, yet, increasing variance significantly.

<table>
<thead>
<tr>
<th>Shape Name</th>
<th>Sample Shape</th>
<th>Test Size</th>
<th>Correct</th>
<th>False</th>
<th>Recognition Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ain_1</td>
<td>200</td>
<td>191</td>
<td>9</td>
<td>95.5%</td>
<td></td>
</tr>
<tr>
<td>Baa</td>
<td>200</td>
<td>193</td>
<td>7</td>
<td>96.5%</td>
<td></td>
</tr>
<tr>
<td>Dal</td>
<td>200</td>
<td>194</td>
<td>6</td>
<td>97%</td>
<td></td>
</tr>
<tr>
<td>Faa</td>
<td>200</td>
<td>187</td>
<td>13</td>
<td>93.5%</td>
<td></td>
</tr>
<tr>
<td>Haa_1</td>
<td>200</td>
<td>196</td>
<td>4</td>
<td>98%</td>
<td></td>
</tr>
<tr>
<td>Thah</td>
<td>200</td>
<td>180</td>
<td>20</td>
<td>90%</td>
<td></td>
</tr>
<tr>
<td>Hlah_1</td>
<td>200</td>
<td>178</td>
<td>22</td>
<td>89%</td>
<td></td>
</tr>
<tr>
<td>Ain_2</td>
<td>200</td>
<td>190</td>
<td>10</td>
<td>95%</td>
<td></td>
</tr>
<tr>
<td>Meem</td>
<td>200</td>
<td>182</td>
<td>18</td>
<td>91%</td>
<td></td>
</tr>
<tr>
<td>Seen</td>
<td>200</td>
<td>175</td>
<td>25</td>
<td>87.5%</td>
<td></td>
</tr>
<tr>
<td>Yaa</td>
<td>200</td>
<td>189</td>
<td>11</td>
<td>94.5%</td>
<td></td>
</tr>
<tr>
<td>Haa_2</td>
<td>200</td>
<td>193</td>
<td>7</td>
<td>96.5%</td>
<td></td>
</tr>
<tr>
<td>Waw</td>
<td>200</td>
<td>183</td>
<td>17</td>
<td>91.5%</td>
<td></td>
</tr>
<tr>
<td>Ain_2</td>
<td>200</td>
<td>195</td>
<td>5</td>
<td>97.5%</td>
<td></td>
</tr>
<tr>
<td>Hlah_2</td>
<td>200</td>
<td>172</td>
<td>28</td>
<td>86%</td>
<td></td>
</tr>
<tr>
<td>Kaf_1</td>
<td>200</td>
<td>196</td>
<td>4</td>
<td>98%</td>
<td></td>
</tr>
<tr>
<td>Lam_1</td>
<td>200</td>
<td>193</td>
<td>7</td>
<td>96.5%</td>
<td></td>
</tr>
<tr>
<td>Lam_2</td>
<td>200</td>
<td>181</td>
<td>19</td>
<td>90.5%</td>
<td></td>
</tr>
<tr>
<td>Raa</td>
<td>200</td>
<td>196</td>
<td>4</td>
<td>98%</td>
<td></td>
</tr>
<tr>
<td>Ssad</td>
<td>200</td>
<td>176</td>
<td>24</td>
<td>88%</td>
<td></td>
</tr>
<tr>
<td>Kaf_2</td>
<td>200</td>
<td>192</td>
<td>8</td>
<td>96%</td>
<td></td>
</tr>
<tr>
<td>Noon</td>
<td>200</td>
<td>196</td>
<td>4</td>
<td>98%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>22 classes</td>
<td>4400</td>
<td>4400</td>
<td>6.18%</td>
<td>93.82%</td>
</tr>
</tbody>
</table>

VI. CONCLUSION
In this paper, we have showed how Regression Curves can
be used to model the variation of Handwritten Arabic characters. The 2nd order of Polynomials curves were injected along the skeleton of the proposed shape under study, where the appropriate coefficients which describe the shape were extracted. We then have used the Apparatus of the Expectation Maximization Algorithm to train the set of extracted coefficients within a probabilistic framework to capture the optimal shape variations coefficients. The set of best fitted parameters were then projected in order to recognize handwritten shapes using Bayes rule of factorization. The proposed approach has been evaluated on sets of Handwritten Arabic Shapes for multiple different writers that we have achieved a recognition rate of nearly 94% on corrected registered shape classes.

VII. FUTURE WORK

In this research, there are some shortcomings to the approach. One of which is that the is missing thoroughly comparison between the proposed method and conventional method to demonstrate the effectiveness of such complicated method. The second is that the extracted parameters from the training stage have not been utilized for the purpose of recognition stage in a statistical framework. Thirdly, we further investigate how we use deformation parameters to fit the align curves to such sample shape, specifically when noise is present.

REFERENCES


Abdullah A. AlShaher has received his Bachelor degree in Computer Science from Huston Tillotson University Austin Texas, his Master degree in Computer Science from University of Denver, Denver Colorado, his Ph.D. in Computer Science form University of York UK. Currently taken the position of the head of Computer and Information Systems department in the Public Authority for Applied Education and Training. His main interests is Shape Recognition, Handwritten Character analysis, Optimization, Computer Vision, Pattern Recognition, and Signal Processing.