MIXTURES OF REGRESSION CURVE MODELS FOR ARABIC CHARACTER RECOGNITION

Abdullah A. Al-Shaher

Department of Computer and Information Systems
College of Business Studies, Public Authority for
Applied Education and Training, Kuwait

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 proceed then, 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.

1. INTRODUCTION

Shape recognition has been the focus of many researchers for the past seven 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 a high degree in 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 need to estimate the parameters which describe the data that is encapsulated within a mixture of shapes.

The literature demonstrates many statistical and structural approaches to various algorithms to model shape variations using supervised and unsupervised learning [12] algorithms. Precisely, the powerful Expectation Maximization Algorithm of Dempster [13] has widely been used for such cases. The EM Algorithm revolves around two step procedures. The expectation E step revolves around estimating the parameters of a log-likelihood function and passes 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 come to unchanged. For instance, Jojic and Frey [14] have used the EM algorithm to fit mixture models to the appearances 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. In general, 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 widely used in research by the computer vision society [1] [2] [3] [4] [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 best fits to a series of landmark points. A related topic is a regression analysis which stresses on probabilistic conclusion on how uncertainty occurs when fitting a curve to a set of data landmarks with marginal errors. Regression curves are applied in data visualization [12][13] to capture the values of a function with missing data [14] and to gain relationship between multiple variables.

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

2. REGRESSION CURVES

We would like to extract the best fit modes that describe the shapes under study, hence, multiple image shapes are required and explained through training sets of class shape ω and 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

$$X^{\omega} = \left((x_1^{\omega_1}, y_1^{\omega_1}), \dots, (x_k^{\omega_1}, y_k^{\omega_1}), (x_1^{\omega_2}, y_1^{\omega_2}), \dots, (x_k^{\omega_2}, y_k^{\omega_2}), (x_1^{\omega_T}, y_1^{\omega_T}), \dots, (x_k^{\omega_T}, y_k^{\omega_T}) \right) \quad (1)$$

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

$$f(x_k^{\omega_T}) = a_0 + a_1 x_k^{\omega_T} + a_2 (x_k^{\omega_T})^2 + a_3 (x_k^{\omega_T})^3 + \cdots + a_j (x_k^{\omega_T})^j = a_0 + \sum_{\tau=1}^{j} a_x (x_k^{\omega_T})^{\tau} \quad (2)$$

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

$$err = \sum (d_i)^2 = (y_1^{\omega_T} - f(x_1^{\omega_1}))^2 + (y_2^{\omega_T} - f(x_2^{\omega_1}))^2 + (y_3^{\omega_T} - f(x_3^{\omega_1}))^2 + (y_4^{\omega_T} - f(x_4^{\omega_1}))^2 \quad (3)$$

where the goal is to minimize the error, we substitute the form of equation (3) with a general least square error

$$err = \sum_{k=1}^{T}(y_k^{\omega_T} - (a_0 + a_1 x_k^{\omega_T} + a_2 x_k^{\omega_T{}^2} + a_3 x_k^{\omega_T{}^3} + \cdots + a_j x_k^{\omega_T{}^j}))^2 \qquad (4)$$

where T is the number of pattern set, k is the current data landmark point being summed, j is the order of polynomial equation. Rewriting equation (4) in a more readable format

$$err = \sum_{k=1}^{T}(y_k^{\omega_T} - (a_0 + \sum_{k=1}^{j} a_k x^k))^2 \qquad (5)$$

Finding the best fitting curve is equivalent to minimize the squared distance between the curve and landmark points. The aim here is to find the coefficients, hence, solving the equations by taking the partial derivative with respect each coefficient a$_0$, .. ,a$_k$; for $k = 1 .. j$ and set each to zero in

$$\frac{\partial err}{\partial a_0} = \sum_{k=1}^{T}(y_k^{\omega_T} - (a_0 + \sum_{K=1}^{T} a_k x^k)) = 0 \qquad (6)$$

$$\frac{\partial err}{\partial a_1} = \sum_{k=1}^{T}(y_k^{\omega_T} - (a_0 + \sum_{K=1}^{T} a_k x^k))x_k^{\omega_T} = 0 \qquad (7)$$

$$\frac{\partial err}{\partial a_2} = \sum_{k=1}^{T}(y_k^{\omega_T} - (a_0 + \sum_{K=1}^{T} a_k x^k))x_k^{\omega_T{}^2} = 0 \qquad (8)$$

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

$$\begin{bmatrix} T & \sum_{k=1}^{T} x_k^{\omega_T} & \sum_{k=1}^{T} x_k^{\omega_T{}^2} \\ \sum_{k=1}^{T} x_k^{\omega_T} & \sum_{k=1}^{T} x_k^{\omega_T{}^2} & \sum_{k=1}^{T} x_k^{\omega_T{}^3} \\ \sum_{k=1}^{T} x_k^{\omega_T{}^2} & \sum_{k=1}^{T} x_k^{\omega_T{}^3} & \sum_{k=1}^{T} x_k^{\omega_T{}^4} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{T} y_k^{\omega_T} \\ \sum_{k=1}^{T} x_k^{\omega_T} y_k^{\omega_T} \\ \sum_{k=1}^{T} x_k^{\omega_T{}^2} y_k^{\omega_T} \end{bmatrix} \qquad (9)$$

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

$$Ax = B \qquad (10)$$

where

$$A = \begin{bmatrix} T & \sum_{k=1}^{T} x_k^{\omega T} & \sum_{k=1}^{T} x_k^{\omega T\,2} \\ \sum_{k=1}^{T} x_k^{\omega T} & \sum_{k=1}^{T} x_k^{\omega T\,2} & \sum_{k=1}^{T} x_k^{\omega T\,3} \\ \sum_{k=1}^{T} x_k^{\omega T\,2} & \sum_{k=1}^{T} x_k^{\omega T\,3} & \sum_{k=1}^{T} x_k^{\omega T\,4} \end{bmatrix},\ X = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix},\ B = \begin{bmatrix} \sum_{k=1}^{T} y_k^{\omega T} \\ \sum_{k=1}^{T} x_k^{\omega T} y_k^{\omega T} \\ \sum_{k=1}^{T} x_k^{\omega T\,2} y_k^{\omega T} \end{bmatrix} \tag{11}$$

solving X to find the coefficients A, B in

$$X = A^{-1} * B \tag{12}$$

The outcome would be the coefficients a_0, a_1, a_2. We follow the similar procedure to find the coefficient sets of the remaining landmark points.

3. LEARNING REGRESSION CURVES

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

$$\alpha_t = (\, a_{t_{1_1}}, a_{t_{1_2}}, a_{t_{1_3}}, a_{t_{2_1}}, a_{t_{2_2}}, a_{t_{2_3}}, \dots, a_{t_{i_n}} a_{t_{i_n}}) \tag{13}$$

The mean vector of coefficient patterns is represented by

$$\mu = \frac{1}{T} \sum_{t=1}^{T} \alpha_t \tag{14}$$

The covariance matrix is then constructed by

$$\sum = \frac{1}{T} \sum_{t=1}^{T} (\alpha_t - \mu)\, (\alpha_t - \mu)^T \tag{15}$$

The following approach is based on fitting a Gaussian mixture models to a set of training examples of curve coefficients. We further assume that training patterns are independent from one to another; thus, they are neither flagged nor labelled to any curve class. Each curve class ω belongs to a set of curve classes Ω has its own mean μ and covariance matrix $\sum$. With these component elements. For each curve class, we establish the likelihood function for a set of the curve patterns in

$$p(\alpha_t) = \prod_{t=1}^{T} \sum_{w=1}^{\Omega} p(\alpha_t \,|\mu_w, \Sigma_w) \tag{16}$$

Where the term $p(\alpha_t|\mu_w, \Sigma_w)$ is the probability of drawing curve pattern α_t from the curve-class ω. Associating the above likelihood function with the Expectation Maximization Algorithm, the likelihood function can be a process reformed in two steps. The process revolves around estimating the expected log-likelihood function iteratively in

$$q_L\left(C^{(n+1)}\big|C^{(n)}\right) = \sum_{t=1}^{T} \sum_{w=1}^{\Omega} P(\alpha_t, \mu_w^{(n)}, \Sigma_w^{(n)}) \; X \ln \; p(\alpha_t|\mu_w^{(n+1)}, \Sigma_w^{(n+1)}) \tag{17}$$

Where the quantity and $\mu_w^{(n)}$ and $\Sigma_w^{(n)}$ are the estimated mean curve vector and variance covariance matrix both at iteration (n) of the Algorithm. The quantity $p(\alpha_t, \mu_w^{(n)}, \Sigma_w^{(n)})$ is the *a posteriori* probability that the training pattern curve belongs to the curve-class ω at iteration n of the Algorithm. The term $p(\alpha_t|\mu_w^{(n+1)}, \Sigma_w^{(n+1)})$ is the probability of distribution of curve-pattern α_t belongs to curve-class ω at iteration (n + 1)of the algorithm; thus, the probability density is associated with the curve- patterns α_t for (t = 1 … T) to curve-class ω are estimated by the updated construction of the mean-vector $\mu_w^{(n+1)}$, and covariance matrix $\Sigma_w^{(n+1)}$ at iteration n+1 of the algorithm. According to the EM algorithm, the expected log-likelihood function is implemented in a two iterative processes. In the M or maximization step of the algorithm, our aim is to maximize the curve mean-vector $\mu_w^{(n+1)}$, and covariance matrix $\Sigma_w^{(n+1)}$, while, in the E or expectation step, the aim is to estimate the distribution of 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 new estimate is computed by

$$p\left(\propto_t, \mu_w^{(n)}, \sum{}_w^{(n)}\right) = \frac{p(\alpha_t|\mu_w^{(n)}, \Sigma_w^{(n)}) \, \pi_w^{(n)}}{\sum_{w=1}^{\Omega} p(\alpha_t|\mu_w^{(n)}, \Sigma_w^{(n)}) \, \pi_w^{(n)}} \tag{18}$$

where the revised curve-class ω mixing proportions $\pi_w^{(n+1)}$ at iteration *(n + 1)*is computed by

$$\pi_w^{(n+1)} = \frac{1}{T} \sum_{t=1}^{T} p(\alpha_t|\mu_w^{(n)}, \Sigma_w^{(n)}) \tag{19}$$

With that at hand, the distributed curve-pattern α_t to the class-curve ω is Gaussian distribution and is classified according to

$$p\left(\alpha_t\Big|\mu_w^{(n)}, \sum{}_w^{(n)}\right) = \frac{1}{(2\pi)^L \sqrt{|\Sigma_w^{(n)}|}} \exp\left[-\frac{1}{2}(\alpha_t - \mu_w^{(n)})^T X \left(\sum{}_w^{(n)}\right)^{-1} X \,(\alpha_t - \mu_w^{(n)})\right] \tag{20}$$

In the *M*, or Maximization step, our goal is to maximize the curve-class ω parameters. The updated curve mean-vector $\mu_w^{(n+1)}$ estimate is computed using the following

$$\mu_w^{(n+1)} = \sum_{t=1}^{T} p(\alpha_t, \mu_w^{(n)}, \Sigma_w^{(n)})\alpha_t \tag{21}$$

And the new estimate of the curve-class covariance matrix is weighted by

$$\Sigma_w^{(n+1)} = \sum_{t=1}^{T} p\left(\alpha_t, \mu_w^{(n)}, \sum{}_w^{(n)}\right) X(\alpha_t - \mu_w^{(n)})\left(\alpha_t - \mu_w^{(n)}\right)^T \tag{22}$$

Both E, and M steps are iteratively converged, the outcome of the learning stage is a set of curve-class ω parameters such as $\mu_w^{(n)}$ and $\Sigma_w^{(n)}$, hence the complete set of all curve-class Ω are computed and ready to be used for recognition.

With the stroke and shape point distribution models to hand, our recognition method proceeds in a hierarchical manner.

4. RECOGNITION

In this stage, we focus on utilizing the parameters extracted from the learning phase to obtain in shape recognition. Here, we assume that the testing shapes

$$f(t) = \sum_{t=1}^{X} \sum_{i=1}^{n} (x_{t_i}, y_{t_i}), where\ (i = 1..n), (t = 1..X) \tag{23}$$

Hence, each testing pattern is represented by

$$\chi_t = (\ (x_{t_1}, y_{t_1}), (x_{t_2}, y_{t_2}), ... (x_{t_i}, y_{t_i})\ )\ \text{for } (t = 1 .. X\ ) \tag{24}$$

Such testing patterns are classified according to computing the new point position of the testing data χ after projecting the sequence of curve-coefficients by

$$f(x, y) = \sum_{i=1}^{n} (\chi_{t_{y_i}} - (\ \alpha_{t_{i1}}\chi_{t_{x_i}}^2 + \alpha_{t_{i2}}\chi_{t_{x_i}} + \alpha_{t_{i3}})) \tag{25}$$

So the sample shape χ_t is registered to class ω which has the highest probability using Bayes rule over the total curve-classes Ω in

$$\arg\min \frac{f(x, y)}{\sum_{w=1}^{\Omega} f(x, y)} \tag{26}$$

5. 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 examined the approach with 1840 handwritten Arabic character shape patterns for training and 4600 patterns for recognition phase. Figures 1illustrates some training patterns used in this paper. Figure 2 demonstrates single shapes and their landmarks representation.
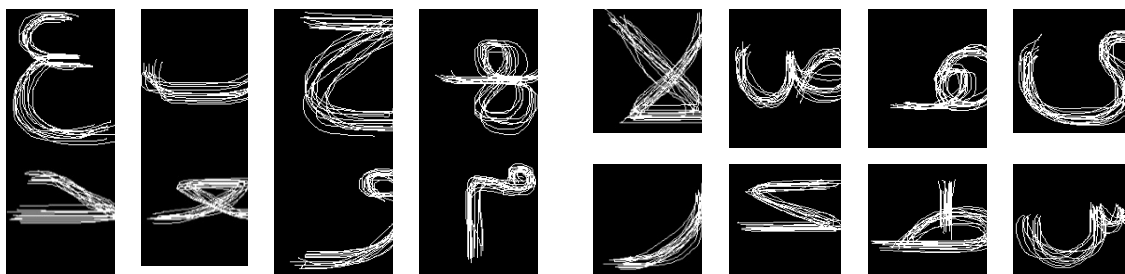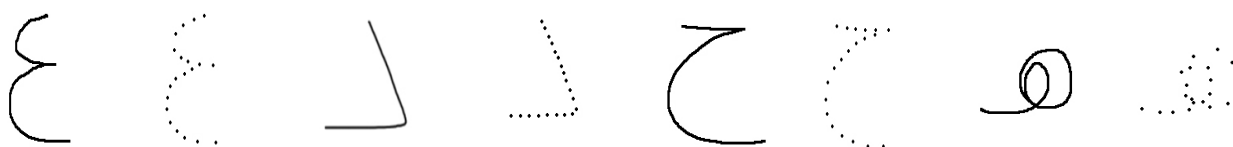

Figure 1: Training sets sample

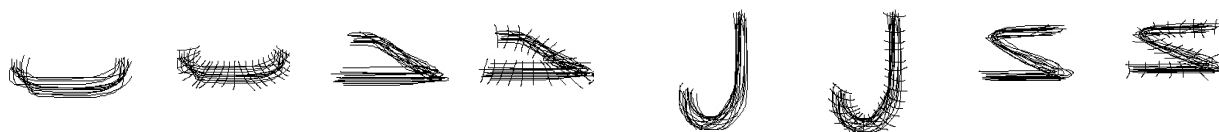**Figure 2:** Training patterns and extracted landmarks

Figure 3: Sample visual of regression Curve-classes

Figure 3indicates regression sample curve-classes as a result of the training stage. Figure 4 demonstrates the curve-classes Ω convergence rate graph as a function per iteration no. in the training phase. The graphs shows how associated distributed probabilities for the set of curve-classes Ω converged into a few iterations.

To take this investigation further, we demonstrate how well the approach behaves in the presence of noise. In figure 5, 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 proves that the recognition rate fails to register shapes to their correct classes in a few iterations and it decreases completely when coordinates are moved away, yet, increasing variance significantly.
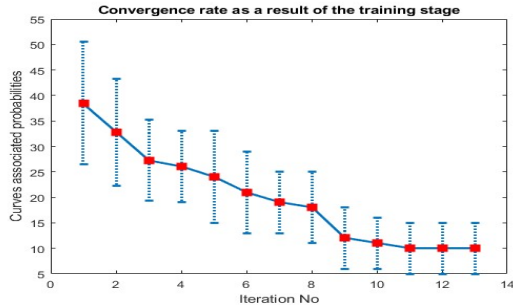
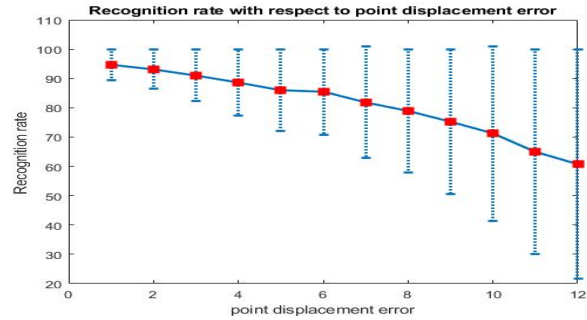**Figure 4:** Convergence Rate as a function per iteration no.

Figure 5: Recognition rate as a function per iteration no with point position error.

Table 1 shows recognition rates per curve-classes ω. Table 1 demonstrates recognition rates per curve-class. In total, we have achieved 94% recognition rate with this approach.

Table 1: Recognition Rates for sample shapes

Sample Shape	Test Size	Correct	False	Recognition Rate	Sample Shape	Test Size	Correct	False	Recognition Rate
ε	200	191	9	95.5%	ڡ	200	176	24	88%
ں	200	193	7	96.5%	ش	200	175	25	87.5%
ٯ	200	183	17	91.5%	؟	200	172	28	86%
ڡ	200	187	13	93.5%	ﻳ	200	181	19	90.5%
ح	200	196	4	98%	ﯾ	200	190	10	95%
ط	200	180	20	90%	م	200	182	18	91%
٥	200	178	22	89%	ٮ	200	193	7	96.5%

6. CONCLUSION

In this paper, we have proved how Regression Curves can be utilized to model the variation of Handwritten Arabic characters. A 2nd order of Polynomials curves are injected along the skeleton of the proposed shape under study, where the appropriate set of curve-coefficients which describe the shape were extracted. We, then have used the Apparatus of the Expectation Maximization Algorithm to train the set of extracted set of curve-coefficients within a probabilistic framework to capture the optimal shape variations coefficients. The set of best fitted parameters are then projected 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 by

which we have achieved a recognition rate of nearly 94% on corrected registered shape classes.

REFERENCES

[1] J. Wood, "Invariant pattern recognition: A review," Pattern Recognition, vol. 29, no. 1, 1996, 1-17.

[2] Anil, K. Jain,Robert P.W. Duin, and Jianchang Mao:"Statistical Pattern Recognition: A Review", IEEE Pattern Analysis and Machine Intelligence, vol 22, No. 1, PP 4-37, 2000.

[3] P.F. Baldi and K. Hornik, "Learning in linear neural networks: A survey," IEEE Transactions on Neural Networks, vol. 6, no. 4, 1995, 837-858.

[4] T.Y. Kong and A. Rosenfeld, "Digital topology: introduction and survey," Computer Vision, Graphics, and Image Processing, vol. 48, no. 3, pp. 357-393, 1989.

[5] T.R. Reed and J.M.H. Dubuf, "A review of recent texture segmentation and feature extraction techniques," CVGIP - Image Understanding, vol. 57, no. 3, pp. 359-372, 1993.

[6] S. Sarkar and K.L. Boyer, "Perceptual organization in computer vision - a review and a proposal for a classifactory structure", IEEE Transactions on Systems Man and Cybernetics, vol. 23, no. 2, 1993, 382-399.

[7] Lorigo L. M., Govindaraju V., "Offline Arabic Handwriting Recognition: A Survey", IEEE Trans Pattern Anal Mach Intelligence, Vol: 28(5): PP 712-24, 2006.

[8] Ishani Patel, ViragJagtap, Ompriya Kale, "A Survey on Feature Extraction Methods for Handwritten Digits Recognition", International Journal of Computer Applications, Vol107, No12, PP 11-17, 2014.

[9] Muhammad Sharif, Farah Naz, Mussarat Yasmin, Muhammad AlyasShahid, AmjadRehman, "Face Recognition: A Survey", Journal of Engineering Science and Technology Review, Vol 10, No 2, PP 166-177, 2017.

[10] Rafiqul Zaman Khan, Noor Adnan Ibraheem, "HAND GESTURE RECOGNITION: A LITERATURE REVIEW ",International Journal of Artificial Intelligence & Applications, Vol.3, No.4, pp 161-174, 2012.

[11] Bruce George Lindsay, Mary L. Lesperance , "A review of semiparametric mixture models", Journal of Statistical Planning and Inference, Vol: 47 , No: 1 PP 29-39 , 1995.

[12] Christopher M. Bishop, "Neural Networks for Pattern Recognition", Clarendon Press, 1995, ISSN 0198538642.

[13] A. Dempster, N. Laird, D. Rubin, Maximum likelihood from incomplete data via the em algorithm, J. Roy. Statist. Soc. Ser. 39 (1977) 1–38

[14] J. Brendan, N. Jojic, Estimating mixture models of images and inferring spatial transformations using the em algorithm, IEEE Comput. Vision Pattern Recognition 2 (1999) 416–422.

[15] C. Bishop, J. Winn, Non-linear Bayesian image modelling, Proceedings of Sixth European

[16] N. Vasconcelos, A. Lippman, A probabilistic architecture for content-based image retrieval, Proceedings of International Conference on Computer Vision and Pattern Recognition, 2000, pp. 216–221.

[17] B. North, A. Blake, Using expectation-maximisation to learn dynamical models from visual data, Image Vision Computing. 17 (8) (1999) 611–616.

[18] M. Revow, C. Williams, G.E. Hinton, Using generative models for handwritten digit recognition, IEEE Trans. Pattern Anal. Mach. Intell. 20 (2) (1996) 592–606

[19] Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer Science, and Business Media, 2006.

[20] T. Cootes, C. Taylor, A mixture model for representing shape variations. Image and Vision Computing, 17(1999) 403-409

[21] AlShaher Abdullah, Hancock Edwin, Learning mixtures of Point Distribution models with the EM algorithm. Pattern Recognition, 36(2003) 2805-2818.