Development of Online Telugu Character Recognizer to recognize even isolated Telugu characters

Josyula Siva Phanriam
Research Scholar ,Krishna University, Machilipatnam, Krishna Dist. Andhra Pradesh.

Dr. G. P. Saradhi Varma
Principal, S.R.K.R Engineering College, Bhimavaram, West Godavari, Andhra Pradesh

Dr. Y.K. Sundra Krishna
Principal, Krishna University, Machilipatnam.

Abstract: - The article brings into limelight the importance of online language recognizing software at present, the author had referred to several distinguished journals and magazines to understand the scope for development of such a tool The author understand the need for development of an online Telugu Character recognition software. The advent of speech interface software, digital ink based writing software brought a revolutionary change in the human and computer interaction. These devices have made keying of information and inputs of information easy when compared with keying information using a key board. One of the advantages with use of these devices and software is that they are noise free and does not cause any ergonomic complications. Thus, the study calls for the development of a online Telugu Character recognition software which can recognize even remotely used Telugu Characters. The software is developed by using Sobolev Coefficients which contains ordered sequences of coordinates created by using Vector support machines. The model is an efficient one when compared with other models in use because it can save time and recognition performance. The higher performance is possible because of the dimensionality of the vector, below 30 in this case.

Keywords- Online recognizer, Telugu handwriting recognition, handwritten character recognition, multi-classifier based recognition, support vector machines, Legendre Sobolev coefficients.

I. INTRODUCTION

One of the latest technological advancements is the development of recognition software, it can be used for recognizing voice, writing purpose and for penning signatures and other customized writing purposes. These software literally replaced the need for keying through keyboard. Also, one of the advantages using these recognition software is the ability to create special symbols.

Telugu language for example, needs at least 3-4 key strokes to produce some particular symbols, it is one of the most popular languages used in the world. We know that hand writing is a way to express our opinions and thoughts. Handwriting interfaces provides better advantages by creating lesser noise, better interaction and ability to create difficult characters. These gadgets ensure privacy, they have the ability to recognize characters when they are written on the device surface.

Online handwriting devices took their second birth nearly 35 years ago[1] although they are born nearly 60 years ago [2]. Despite the fact that many such cost effective, high performance commercial systems as well as software packages are available for identifying printed text in languages like English, Roman, Chinese and Japanese besides the Microsoft's transcriber to recognize cursive English to use with PDAs, the handwritten script recognition is strill considered to be in infancy.

The study calls for the development of a complete OCR system for identifying Telugu printed text as software package took place in 2001 [3], some attempts were made nearly forth five years ago Deekshatulu[4]. However, recognition of Telugu handwriting through device interfaces took place nearly eight years back [5] and is the only noticeable effort made so far.

One of the issues hindering the growth of the recognition devices and the software are the difficulty in identification is due to lack of publicly available standardized and validated data sets, the varied and complex structure of Telugu script and the multiple variations in writing style [6]. Recently HP Research Labs, Bangalore produced online and offline data sates [7] which are standardized and also produced hand writing recognition tools [8].

Telugu language is derived from Dravidian language and is termed as a syllabic language and written in phonographic Brahmi script. Both the Dravidian language and the Telugu language use the same writing form and use the same irrespective of variety of dialects exists among the Telugu speaking people. Telugu language consists of 52 basic symbols composed of 16 vowels and 36 consonants. It consists of vowel sound modifying symbols for consonants termed as maatras, and core consonant sound combined with another consonant symbol termed as voththulu, akin to the form of consonant conjuncts. Both the maatras and the voththulu, together are termed as guninthalu. Thus, there are thousands of charcecters to be recognized in Telugu, They also contain the inseparable base consonants and their maatras (ఖి,తౌ), the voththu, which are visually different from base consonant (క్క, క్త, క్న), the inherent similarity between different symbols (థ,ద, ధ), same as well as different maatra positional variations (ము,లు, సా), positional variations of voththulu (క్క, క్త) and the multilevel structure of the symbols (స్త్రౌ) further added to the recognition complexity. Also, the difference in writing habits of the writers made things extremely complex for the development of recognition devices. Some of them write top to bottom, from left to right and in different directions, also one of the distinct features of the Telugu Language is the lack of difference between upper case characters and lower case characters.

The article discusses the literary works referred by the author to complete the article; sections, 3 and 3 are used for discussing the extraction and classifier used for recognition purposes. Section 5 is used for bringing a comparison of the proposed model with the existing model and section 6 is used for concluding the article.

II. PREVIOUS CONTRIBUTORY WORKS

Recognition of Telugu characters is based on decomposition strategies, in the first , the large sets of symbols are adapted and in the second, the constituent strokes and themes are adapted. The division is having several draw backs. [6]. First, these strokes are not known in advance. Second, depends on the script and writing styles. Third, Initially the figure out the unique strokes requires manual intervention. Fourth difficulty lies in finding out the lack of standard procedure to determine unique strokes. Fifth problem lies in finding the necessity for a huge number of rules to build the symbols of language. Finally, require significant manual intervention in the training phase.

Decomposing to constituent graphemes strategy is very effective with constrained handwriting and requires less effort for dataset creation. The dataset created by HP Labs is based on graphemes. This strategy does not address stroke-order variations across symbols in the character, symbol-order variations, co-articulation effects, and the cases where opaque symbols are created by CC and CV combinations [6].

Swethalakshmi et al. [5] developed an online translator for the first time using supported vector machines with pre-processed stroke coordinates as features. The pre-processing of the characters consists of normalization, smoothing and retracing. The number of points fixed as the average number of points for that character in the dataset. The single engine SVM with σ=30 and C = 50 reported a recognition of 83.08%.

Jayaraman et al.[9] developed an approach called as modular approach to online recognition with 86.9% stroke level accuracy. He and his team designed a separate SVM engine to identify base strokes (179), bottom strokes (44) and top strokes (30) totalling to 253 stroke classes. In order to obtain better accuracy he classified the strokes into these 3 groups based on relative position of the stroke in the character.

Babu et al.[10] proposed a model which reads from left-to-right, termed as HMM based recognizer considering 141 out of 161 classes from the HP dataset. During this process of development he eliminated the unused and positional variations of maatras. The combination of time and frequency domain features compared to each of them resulted top-1 accuracy of 91.6% and top-5 accuracy of 98.7%r recognition accuracies.

Dynamic Time Warping (DTW) is a recognizer proposed by Prasanth et al. [11] , In this model he considered 141 classes akin to Babu et al.[11]. The curve uniformly resample to 60 points and normalized both x and y values to [0, 10]. The single stage scheme resulted in an efficiency of 90.6% recognition with 0.166 symbols/sec. The two stage scheme improved the speed of functioning to 3.977 symbols/sec with a slight drop in recognition accuracy (89.77%).

Amit et al.[12] offered a more complete online recognizer for Indic languages and tested with Telugu and Malayalam datasets. The authors used a model called as a generative model of handwriting formation to combine the language and script information. The authors used discriminative stroke classification to deal with similar strokes. The system achieved stroke recognition accuracy of 95.12% and 78% of character recognition accuracy on a total amount of 60, 492 words from 367 writers.

Rajkumar et al.[13] used two different approaches, he used the Ternary Search Tree (TST) and Support Vector Machines (SVM) approach. The study involves use of elaborate multi-classifier architectures. Depending on the stroke vertical position, authors divided the stroke sets into 4 different subclasses and a separate SVM trained on each group. Stroke level recognition is based on a bank of SVMs. The TST with about 19,383 rules and the SVM reported recognition performances of 89.59% and 96.69% at stroke level, 90.55% and 96.42% at character level respectively. The SVM method surpasses some of the recent online Telugu script recognition performances.

In this model the same HP Telugu dataset and 141 classes considered by Babu et al.[10] and Prasanth et al.[11] is taken. The author then divided the classes into groups based on the number of constituent strokes and building independent models for each of the group. The model improved the recognition as well as time efficiencies of the recognition engine. To further improve the efficiency, the author suggested a recognition method without pre-processing, a mandatory step in all proposed methods in the literature and easy to extract features with dimensionality in the order of ten.

III. PROPOSED FEATURE EXTARCTION

The works by Watt et al.[14][15][16][17] inspred the author to conduct a study on development of the said model, the author figured the sequence of coordinate points corresponding to online handwritten isolated Telugu character using the truncated series coefficients. The coefficients can be computed easily by using the orthogonally aligned inner products on a functional basis. Three important series representations are used in handwriting identification of mathematical symbols orthogonal on domain [-1, 1]: Chebyshev, Legendre and Legendre Sobolev.

For the first time, Char and Watt [14] approximated handwritten curve using truncated Chebyshev series of order 10. The weight function $w(t) = \frac{1}{\sqrt{1}}$ used in inner products for computing series coefficients hikes the computational complexity. Watt et al.[15] improved the model of strokes using Legendre series by using weight function $w(t) =$ the model is equally good when compared to that of Chebyshev series and efficient in terms of computation. The Legendre series competently approximated handwritten curves with degree reduced to 18. The same authors, later replaced the Legendre series with truncated Legendre-Sobolev series by using degree equal to 12 [16]. The subsequent paragraphs describe the mathematical preliminaries as well as the computation of Legendre Sobolev series coefficients.

If the sequence of coordinate points having metrics of $\lambda \in [0,$ then $f($ and $g($ are termed as coordinate functions. The possible coordinates for coordinate functions are: time, arc length, and affine arc length. The time coordinate is most commonly used but is not useful for handwriting recognition due to uneven distribution of points.

The most commonly used coordinate is arc length but is invariant only to Euclidean transformations and is not shear invariant. In handwriting horizontal shear is very common compared to vertical shear. The affine arc length measurement copes with shear transformations besides invariant to special affine transformations. Hence the affine arc length is used as a measurement and is defined as

$$AAL(\lambda) = \begin{cases} 0 & \lambda = 0 \\ \sqrt[3]{\begin{vmatrix} X(\lambda-2) & Y(\lambda-2) & 1 \\ X(\lambda-1) & Y(\lambda-1) & 1 \\ X(\lambda) & Y(\lambda) & 1 \end{vmatrix}} & \text{otherwise} \end{cases}$$

A family of functions $F = \{f_i$ referred to as the functional basis if every other function of general class can be called as linear combination of the functions of F as

$$f(t) = \sum_{i=0}^{\infty} C_i f_i(t)$$

Where $\{C_i$ are referred to as coefficients. The functions $\{f_i$ are termed as basic functions. The inner product of two measured functions f and g can be obtained by coupling the product of complex conjugate of f with

g and a suitable weight w with respect to parameter t with appropriate integration boundaries. If both f and g are real-valued functions, then the complex conjugate of f becomes f and thus the inner product becomes

$$\langle f, g\rangle = \int_a^b f(t)g(t)w(t)\,dt$$

The Legendre inner product is defined on [-1, 1] with $w(t) =$ and is defined as

$$\langle f, g\rangle_L = \int_{-1}^{1} f(t)g(t)\,dt$$

If the inner product of the functions f and g equals to zero, then the functions can be arranged in the form of orthogonal formation.. Similarly, incase two functions belongs to a functional basis are orthogonal to each other, then that functional basis is called as orthogonal basis. In this study the monomial functional basis $\{t$ is considered and converted into orthogonal/orthonormal basis by using the Gram-Schmidt process defined in terms of inner products. In case one uses Legendre inner product, the obtained orthogonal functions are termed as Legendre polynomials and are defined as

$$P_n(t) = \frac{1}{2^n n!}\frac{d^n}{dt^n}((t^2-1)^n)$$

These Legendre polynomials satisfy the following property

$$\int_{-1}^{1} P_m(t)P_n(t)\,dt = \begin{cases} 0 & m \neq n \\ \dfrac{2}{2m+1} & m = n \end{cases}$$

The shifted Legendre polynomials are defined as $\tilde{P}_n(t) = P_n(2t -$ and maps from the interval [0, 1] to the interval [-1, 1] and are orthogonal on [0, 1]. The polynomials are generated by using

$$\tilde{P}_n(t) = (-1)^n \sum_{k=0}^{n} \binom{n}{k}\binom{n+k}{n}(-t)^k$$

These shifted Legendre polynomials satisfy the following property.

$$\int_{-1}^{1} \tilde{P}_m(t)\tilde{P}_n(t)\,dt = \begin{cases} 0 & m \neq n \\ \dfrac{1}{2m+1} & m = n \end{cases}$$

Now the inner product of the function f with one of the functions f_j is computed by using one of the orthogonal functional basis of f and it results in

$$\begin{aligned} < f, f_j >_L &= \int_{-1}^{1} f(t)f_j(t)\,dt \\ &= \int_{-1}^{1}\sum_{i=0}^{\infty} C_i f_i(t)\, f_j(t)\,dt \\ &= \sum_{i=0}^{\infty} C_i \int_{-1}^{1} f_i(t) f_j(t)\,dt \end{aligned}$$

$$= \sum_{i=0}^{\infty} C_i < f_i, f_j >_L$$

$$= C_j < f_j, f_j >_L \text{ as } < f_j, f_j >_L = 0 \text{ if } i \neq j$$

Thus the Legendre series coefficient computedfrom the coordinate function as

$$C_j = \frac{< f, f_j >}{< f_j, f_j >}$$

The Legendre polynomials are orthogonal on [-1, 1] but the coordinate functions are defined on [0, L]. Hence replacing f(with $\hat{f}(=$ $f\left((\omega + 1)\right.$ scales down the interval from [0, L] to [-1, 1]. Then the coefficient becomes

$$C_j = \frac{2j+1}{2} \int_{-1}^{1} \hat{f}(\omega) f_j(\omega) d\omega$$

$$= \frac{2j+1}{2} \int_{0}^{L} f(\lambda) f_j\left(2\frac{\lambda}{L} - 1\right) \frac{2}{L} \, d\lambda$$

$$= \frac{2j+1}{L} \int_{0}^{L} f(\lambda) f_j\left(2\frac{\lambda}{L} - 1\right) d\lambda$$

$$= \frac{2j+1}{L} \int_{0}^{L} f(\lambda) f_j(2u - 1) d\lambda, \text{ by putting } u = \frac{\lambda}{L}$$

$$= \frac{2j+1}{L} \int_{0}^{L} f(\lambda) \sum_{k=0}^{j} \left(\text{coefficient of } u^k \text{ in } f_j(2u-1)\right) u^k \, d\lambda$$

$$= \frac{2j+1}{L} \sum_{k=0}^{j} \left(\left(\text{coefficient of } u^k \text{ in } f_j(2u-1)\right) \int_{0}^{L} f(\lambda) u^k \, d\lambda \right)$$

$$= \frac{2j+1}{L} \sum_{k=0}^{j} \left(\left(\text{coefficient of } u^k \text{ in } f_j(2u-1)\right) \int_{0}^{L} f(\lambda) \frac{\lambda^k}{L^k} \, d\lambda \right)$$

$$= \frac{2j+1}{L}\sum_{k=0}^{j}\left(\frac{\left(\text{coefficient of } u^k \text{ in } f_j\left(2\frac{\lambda}{L}-1\right)\right)}{L^k}\int_0^L f(\lambda)\,\lambda^k\ d\lambda\right)$$

By substituting shifted Legendre coefficient from equation and defining numerical moment of the coordinate function as

$$M_k(f,L) = \int_0^L f(\lambda)\lambda^k\ d\lambda$$

Now the Legendre series coefficient equals to

$$C_j = (-1)^j\frac{2j+1}{L}\sum_{k=0}^{j}\left(\frac{-1}{L}\right)^k\binom{j}{k}\binom{j+k}{k}M_k(f,L)$$

Now we compute the numerical moment $M_k(f,$ by applying piecewise integration and thus equal to

$$M_k(f,L) = \sum_{i=0}^{L}\int_i^{i+1} f(\lambda)\lambda^k\ d\lambda$$

The integration $\int_i^{i+1} f(\lambda)\lambda^k$ is approximated as multiplication of the exact integral of on [i, i+1] with the mean value of the function at end points.

$$\int_i^{i+1} f(\lambda)\lambda^k\ d\lambda \approx \frac{(i+1)^{k+1}-i^{k+1}}{k+1}\frac{f(i+1)+f(i)}{2}$$

By substituting the approximation, the numeric moment $M_k(f,L)$

$$= \sum_{i=0}^{L-1}\frac{(i+1)^{k+1}-i^{k+1}}{k+1}\frac{f(i+1)+f(i)}{2}$$

$$= \sum_{i=0}^{L-1}\frac{(i+1)^{k+1}}{k+1}\frac{f(i+1)+f(i)}{2} - \sum_{i=0}^{L-1}\frac{i^{k+1}}{k+1}\frac{f(i+1)+f(i)}{2}$$

$$= \sum_{i=1}^{L}\frac{i^{k+1}}{k+1}\frac{f(i)+f(i-1)}{2} - \sum_{i=0}^{L-1}\frac{i^{k+1}}{k+1}\frac{f(i+1)+f(i)}{2}$$

$$= \sum_{i=1}^{L}\frac{i^{k+1}}{k+1}\frac{f(i)+f(i-1)}{2} - \sum_{i=1}^{L-1}\frac{i^{k+1}}{k+1}\frac{f(i+1)+f(i)}{2}$$

$$= \left(\frac{L^{k+1}}{k+1}\frac{f(L)+f(L-1)}{2}\right) + \sum_{i=1}^{L-1}\frac{i^{k+1}}{k+1}\frac{f(i)+f(i-1)}{2} - \sum_{i=1}^{L-1}\frac{i^{k+1}}{k+1}\ \mathrm{X}\ \frac{f(i+1)+f(i)}{2}$$

$$= \left(\frac{L^{k+1}}{k+1}\frac{f(L)+f(L-1)}{2}\right) + \sum_{i=1}^{L-1}\frac{i^{k+1}}{k+1}\ \mathrm{X}\ \frac{f(i)+f(i-1)-f(i+1)-f(i)}{2}$$

Finally,

$$M_k(f, L) = \left(\frac{L^{k+1}}{k+1}\frac{f(L) + f(L-1)}{2}\right) + \sum_{i=1}^{L-1}\frac{i^{k+1}}{k+1}\frac{f(i-1) - f(i+1)}{2}$$

From the above equation, it can be identified that the numerical moment computation in turn the Legendre series coefficients are computed simultaneously while the curve is being drawn without waiting for the entire curve to be completed. The process helps in reducing the time to compute the coefficients. Thus in the proposed work truncated Legendre Sobolev series coefficients with degree equal to 12 are extracted for the dataset downloaded from HP Research Labs[7] and the detailed steps are summarized as shown below:

- Computation of orthonormal Legendre-Sobolev Polynomial coefficients for a single time whose degree varies from 0 to 12 by using Gram-Schmidt process.
- The multiple strokes comprising the symbol into a single stroke in the order in which they written are joined if necessary. .
- Compute affine arc lengths, the affine invariant parameterization used for curve approximation at each point of the ordered sequence of coordinates corresponding to ink traces[17].
- Compute numerical moment $M_K(f,$ for each of the coordinate functions $X($ and $Y($ where $\lambda \in [0,$ and $K \in [0, 1$. The numerical momentcomputation is overlapped with the writing of the character [15].
- Compute the 13 Legendre Sobolev series coefficients for each of the coordinate functions $X($and $Y($ referred to as $LSX_0, LSX_1, LSX_2, \cdots, LS$and $LSY_0, LSY_1, LSY_2, \cdots, LS$ respectively.
- The pair $(LSX_0, LS$ represents the curve centroid and hence dropped to bring translation invariance to the feature set.
- Arrange the remaining coefficients into 24-dimension feature vector in any order. We chosen $LSX_1, LSY_1, LSX_2, LSY_2, \cdots, LSX_{12}, LS$ order.
- To achieve scale invariance, normalize the feature vector by dividing it with Euclidean norm represented $\|FV$[18].

$$\|FV\|_2 = \sqrt{LSX_1^2 + LSY_1^2 + LSX_2^2 + LSY_2^2 + \cdots + LSX_{12}^2 + LSY_{12}^2}$$

- To achieve storage and computational efficiency the coefficients are scaled and rounded to integer values in the range [-127, 127][15].

To handle the confusing pairs (ఛ§, బ§) which are a result of the difference in writing styles, the author added four features to the existing feature. The features added are as follows: the cosine and sine angles with the horizontal axis by the first point to last point vector and first point to third point vector. Sometimes the first and last point may be on the same line and hence to avoid this problem, the cosine and sine angles are treated as zero if the magnitude of the vector falls below $\frac{\max(\text{width}, \text{heig}}{4}$. In case the vector represented by the two points $(X_1,$ and $(X_2,$then cosine and sine angles are computed by using the formulas shown below:

$$\cos(\theta) = \frac{X_2 - X_1}{\sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}}$$

$$\sin(\theta) = \frac{Y_2 - Y_1}{\sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}}$$

The 24-dimensional Legendre Sobolev coefficient feature vectors together with the 4 dimensional features results in 28-dimensional feature vector. These feature vectors are used in training and testing phases of proposed character recognition

IV. PROPOSED MULTI-CLASSIFIER BASED RECOGNITION METHOD

The HP Labs dataset [7] is a combination of 29188 training samples collected from 111 writers and 9224 test samples, they are collected from 35 writers apart from the writers used for training data. The symbols are divided into five different overlapped groups (class : group = 1:M) based on the number of constituent strokes. By considering possible ways of writing each of the character belongs to 141 classes, an interval is defined consisting of the minimum and maximum number of constituent strokes are fixed. The samples whose stroke count lies outside the specified range are not considered for the study. Thus the resulting train and test samples are 28482 and 8980 respectively. The summary of stroke count and the number of classes are tabulated in Table 1.

Table 1pre-classification of character classes

Stroke count	Number of classes
1	80
2	84
3	47
4	18
5	05

The training samples are used to build support vector machine models for each group. The parameters used for model training are: cost (C) = 1.0, gamma = 1 / number of classes, third degree polynomial as kernel. The probability estimates are produced as output instead of the default 0/1 output to calculate the top-n recognition rates. The overall recognition performance in addition to the individual performances of the models are tabulated in Table 2.

Table 2 proposed method recognition performance

Model	Samples	Recognized	Recognition (%)
SVM-1	4690	4348	92.71
SVM-2	3083	2734	88.68
SVM-3	997	853	85.56
SVM-4	191	149	78.00
SVM-5	19	17	89.47
	8980	8101	**90.21**

V. RESULT COMPARISON

The overall recognition performances up to top-5 are compared with the methods proposed by Babu et al.[10] and Prashant et al.[11] based on the same dataset and same number of classes are tabulated in Table 3.

Table 3 comparison of proposed method with Babu et al. [10] and Prasanth et al. [11]

	Proposed	Babu et.al. [10]	Prasanth et al. [11]	
			1-stage	2-stage

Top-1	90.21	91.60	90.60	89.77
Top-2	96.43	97.00	-	-
Top-3	97.98	98.00	-	-
Top-4	98.53	98.40	-	-
Top-5	99.00	98.70	-	-

The model proposed by the author displayed better results in terms of performance when compared with the other methods referred by the author. The models displayed better performance compared to Swethalakshmi et al.[5] (83.08%), Jayaraman et al.[9] (86.9%) and Rajkumar et al.[13](89.59%) using ternary search method. All these approaches are based on stroke level and our approach definitely display better performance in stroke level recognition as shown in the Table 2,the highest recognition rates (92.71%) achieved for the model corresponding to single stroke.

When the results obtained in the proposed model are compared to the results obtained by Amit et al.[12](95.12%) and Rajkumar et al [15] (96.69%) using SVM based approach, the proposed method reported less recognition rates. These two methods are completely using different datasets and the HP Telugu dataset is poorly written as stated by Babu et al.[10]. If top-2 and top-3 classes pushed to top-1, the results are on par with the above methods.

VI. CONCLUSION

The model proposed by the author displayed recognition performance which is on par with the methods proposed in the literature with only 28-dimensional sample to compute feature vector compared to 100 of complex features proposed by most of the methods in the literature. The skipping of pre-processing, overlapped feature extraction with the acquisition of online data, pre-classification before recognition and the multi-classifier based recognizer helped to a great extent in reducing the computational complexity.

By analysing the results manually we can understand that a few classification errors took place and they are due to stroke order variations and confusing pairs. To further improve the recognition performance, variant analysis is suggested and creating models corresponding to each variant, addition of more features and/or stages needs to be explored. The robustness of the proposed model can be tested with other standard data sets and the possibility of utilizing the same recognizer for offline static character images by converting them into ordered sequences of coordinates needs to be explored.

REFERENCES

[1] C. T. Charles, Y. S. Ching and W. Toru, "The state of the art in online handwriting recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 12, no. 8, pp. 787-808, August 1990.

[2] A. L. Koerich, R. Sabourin and C. Y. Suen, “Large vocabulary off-line handwriting recognition: A survey,” *Pattern Analysis and Applications,* vol. 6, no. 2, pp. 97-121, 2003.

[3] A. Negi, B. Chakravarthy and K. B, "An OCR System for Telugu," in *Proc. of 6th International Conference on Document Analysis and Recognition (ICDAR 2001)*, Seattle, WA, USA, 2001.

[4] S. N. S. Rajasekaran and Deekshatulu, "Generation Recognition of Printed Telugu Characters," *Computer Graphics and Image Processing,* vol. 6, no. 4, pp. 335-360, 1977.

[5] H. Swethalakshmi, J. Anitha, C. V. Srinivasa and S. C. Chandra, "Online Handwritten Character Recognition of Devanagari and Telugu Characters using Support Vector Machines," in *10th International Workshop on Frontiers in Handwriting Recognition*, La Baule, France, 2006.

[6] G. Venu and S. Srirangaraj, Guide to OCR for Indic Scripts - Document Recognition and Retrieval, London: Springer-Verlag, 2009.

[7] "Telugu Character Dataset," [Online]. Available: http://lipitk.sourceforge.net/datasets/teluguchardata.htm.

[8] "Lipi Toolkit Project," [Online]. Available: http://lipitk.sourceforge.net/.

[9] A. Jayaraman, S. C. Chandra and C. V. S, "Modular Approach to Recognition of Strokes in Telugu Script," in *Ninth International Conference on Document Analysis and Recognition (ICDAR)*, Parana, 2007.

[10] V. Babu, L. Prasanth, R. Sharma, G. V. Rao and A. Bharath, "HMM-Based Online Handwriting Recognition System for Telugu Symbols," in *Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR)*, Chicago, 2007.

[11] L. Prasanth, V. J. Babu, R. R. Sharma, G. V. P. Rao and M. Dinesh, "Elastic Matching of Online Handwritten Tamil and Telugu Scripts Using Local Features," in *Proceedings of Ninth International Conference on Document Analysis and Recognition (ICDAR)*, Curitiba, Brazil, 2007.

[12] A. Amit and M. N. Anoop, "A Hybrid Model for Recognition of Online Handwriting in Indian Scripts," in *International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Kolkata, 2010.

[13] J. Rajkumar, K. Mariraja, K. Kanakapriya, S. Nishanthini and V. S. Chakravarthy, "Two Schemas for Online Character Recognition of Telugu Script Based on Support Vector Machines," in *International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Bari, 2012.

[14] " Char, B. W. and Watt, S. M., "Representing and characterizing handwritten mathematical symbols through succinct functional approximation,"," in *Proc. of International Conference on Document Analysis and Recognition*, 2007.

[15] O. Golubitsky and S. M. Watt, "Online stroke modeling for handwriting recognition," in *Proc. Center for Advanced Studies (CASCON '08)*, New York, NY, USA, 2008.

[16] O. Golubitsky and S. M. Watt, "Online Computation of Similarity between Handwritten Characters," in *Proc. Document Recognition and Retrieval XVI (DRR 2009)*, 2009.

[17] O. Golubitsky, V. Mazalov and S. M. Watt, "Toward affine recognition of handwritten mathematical characters," in *Proc. of International Workshop on Document Analysis Systems, (DAS 2010)*, Boston, USA, 2010.

[18] O. Golubitsky and S. M. Watt, "Distance-based classification of handwritten symbols.," *International Journal of Document Analysis and Recognition ,* 2010.