International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 02 | Feb 2020 www.irjet.net p-ISSN: 2395-0072

CONVERSION OF ANCIENT TAMIL CHARACTERS TO MODERN TAMIL CHARACTERS

Dhivya S[1], Supriya R[2]

[1]Assistant Professor, Department of Computer Science and Engineering, Jeppiaar SRR Engineering college.
[2]UG Student, Department of Computer Science and Engineering, Jeppiaar SRR Engineering college, Chennai, Tamil Nadu, India

---***---

Abstract- *The conversion of Ancient Tamil characters to trendy text could be a necessary automation in image process. These ancient Tamil characters square measure pictures taken from ancient Tamil stone encryption and recognizing and traditional Tamil characters could be a powerful task for modern generation who learn to browse and write solely with modern Tamil characters. Learning the evolution of contemporary Tamil from ancient Tamil is time overwhelming method so a recognition system helps to show, perceive and conjointly to analysis the traditional cultures and heritages. To design a good recognition system, we have a tendency to propose a technique known as noise removal that is additionally known as pre-processing that removes the whole disturbance within the input image. A technique known as morphological operation to perform dilation and erosion operations and therefore the connected part to seek out the letters that square measure gift within the binary image. Finally, we are going to phase every and each character and match it with our current Tamil language employing a methodology known as 'corpus analysis' and can turn out the matching letters as the result.*

1. INTRODUCTION

Tamil character recognition has always been an active field of research for computer scientists worldwide due to its useful real-life applications such as automatic data entry, mail processing and form processing Character recognition is a classic problem in the field of image processing and neural networks. The script used by these inscriptions is commonly known as the Tamil script, and differs in many ways from standard Ashokan Brahmi. For example, early Tamil Brahmi, unlike Asoka Brahmi, had a system to distinguish between pure consonants and consonants with an inherent vowel.

Vatteluttu alphabet is orthography originating from the normal Tamil people of Southern India. Developed from the Tamil (Tamil-Brahmi), Vatteluttu is one among the alphabet systems developed by Tamil people to write down the Proto-Tamil language. it's currently spoken by about 77 million people round the world with 68 million speakers residing in India mostly within the state of Tamil Nadu. It's one among the official languages in India, Sri Lanka and Singapore. Tamil characters contains small circles or loops, which are difficult to acknowledge. Recognizing the traditional Tamil characters that's, the Vatteluttu alphabets may be a tough task for the fashionable generation who learn to read and write only with modern Tamil characters. Learning the evolution of recent Tamil from ancient Tamil may be a time-consuming process therefore a recognition system helps to show, understand and also to research the traditional cultures and heritages. to style an honest recognition system this paper proposes feature extraction of acquired images.

Handwritten character recognition is one among the foremost difficult tasks within the pattern recognition system. There are lot of difficult things need in many image processing techniques to solve. The difficulties are, how to separate cursive characters into an individual character, how to recognize unlimited character fonts and written styles, and how to distinguish characters that have the same shape but different meaning such as character „o‟ and number„0‟. Many researchers plan to apply many techniques for breaking through the complex problems of handwritten character recognition. There are many applications need to take advantage of the handwritten character recognition system namely, automatic reading machine, non-keyboard computer system, and automatic.

2. RELATED WORKS

Mingli Zhang and Christian Desrosiers [1] "High-quality Image Restoration using Low-Rank Patch Regularization and Global Structure Sparsity" -IEEE Transactions on image processing, vol: 28, no: 2, oct 2018. In recent years, approaches based on nonlocal self-similarity and global structure regularization have led to significant improvements in image restoration. Nonlocal self-similarity exploits the repetitiveness of small image patches as a powerful prior in the reconstruction process. Likewise, global structure regularization is based on the principle that the structure of objects in the image is represented by a relatively small portion of pixels. Enforcing this structural information to be sparse can thus reduce the occurrence of reconstruction artefacts. So far, most image restoration approaches have considered one of these two strategies, but not both. This paper presents a novel image restoration method that combines nonlocal self-similarity and global structure sparsity in a single efficient model. Group of similar patches are reconstructed simultaneously, via an adaptive regularization technique based on the weighted nuclear norm. Moreover, global structure is preserved using an innovative strategy, which decomposes the image into a smooth component and a sparse residual, the latter regularized using l_{1} norm. An optimization technique, based on the alternating direction method of multipliers algorithm, is used to recover corrupted images efficiently. The performance of the proposed method is evaluated on two important image restoration tasks: image completion and super-resolution. Experimental results show our method to outperform state-of-the-art approaches for these tasks, for various types and levels of image corruption.

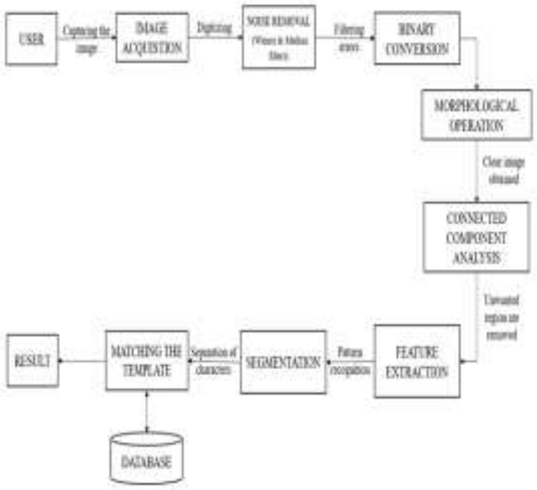
Po-Hsiung Lin, Bo -Hao Chen, Fan-Chieh Cheng, Shih-Chia Huang [2]," A Morphological Mean Filter for Impulse Noise Removal ", Journal of Display Technology, vol: 12, April 2016. Median filtering computation for noise removal is often used in impulse noise removal techniques, but the difficulties in removing high-density noise aspect restrict its development. In this paper, we propose a very efficient method to restore image corrupted by high-density impulse noise. First, the proposed method detects both the number and position of the noise-free pixels in the image. Next, the dilatation operation of the noise-free pixels based on morphological image processing is iteratively executed to replace the neighbour noise pixels until convergence. By doing so, the proposed method is capable to remove high-density noise and therefore reconstruct the noise-free image. Experimental results indicate that the proposed method more effectively removes high-density impulse noise in corrupted images in comparison with the other tested state-of-the-art methods. Additionally, the proposed method only requires moderate execution time to achieve optimal impulse noise removal.

G. Bhuvaneswari, and V. Subbiah Bharathi [3], "An efficient method for digital imaging of ancient stone inscriptions", current science, vol: 110, no: 2, Jan. 2016.Ancient stone inscription is one of the most important primary sources to know about our ancient world such as age, art, politics, religion, medicine, etc. Image acquisition is the first stage for digitizing and preserving the stone inscriptions for further reference. The traditional method of wet paper squeezes is still being used, that will be digitized and preserved for recognition. In this communication, we propose a new image acquisition method called shadow photometric stereo method for upgrading the image for recognition. The efficiency of the proposed acquisition method has been proved in image thinning process. Improving the thinning quality of the characters facilitates better feature extraction for character recognition. An experiment has been performed on two stone inscriptions that were in different places, one inside laboratory and other in its original place, i.e. outside the laboratory. Analyses were performed in terms of performance measures such as hamming distance and peak signal-to-noise ratio. Comparisons with the best available results are given to illustrate the best possible technique that can be used as a powerful image acquisition method.

3. PROPOSED SYSTEM

Here in our proposed system first we have to collect the database of characters those belongs to a different century. In computer vision, and machine recognition of patterns the need for reducing the amount of information to be processed to the minimum is necessary. The thinned characters are used for recognition. In addition, the reduction of a picture to its essentials can eliminate some contour distortions while retaining significant topological and geometric properties. The pre-processing step dedicates to acquiring the image from stone inscription. The recognition step involves training the classifier and then testing a new input based on the trained data.

4. DATAFLOW DIAGRAM



5. METHODOLOGY

5.1 IMAGE ACQUISITION AND NOISE REMOVAL (PREPROCESSING):

Image acquisition is that the first stage for digitizing and preserving the stone inscriptions for further reference. The traditional method of wet paper squeezes remains getting used , which can be digitized and preserved for recognition. Noise is that the results of errors within the image acquisition process that end in pixel values that don't reflect truth intensities of the important scene. Noises result in distortions of final result so it has to be removed. The Median filtering Technique is used for removing noise and Wiener filter to remove the 'Gaussian' noise. The input image of the Tamil historical inscriptions could also be degraded because of the presence of the broken characters, erased characters, touching characters, distortion because of fossils settled, irrelevant symbols engraved by the scribes and so on. The non-uniform spacing between the lines and characters of epigraphically images and the skew could complicate the process of deciphering the script. Hence, Median filter technique is adopted for removing noise from the scripted image. The process is estimated using the following equation.

□□ =□ □□ □ =□−□□ (1) □ □□□□□□□□□

5.2 BINARY CONVERSION:

Binarization is the process of converting a grey scale image (0 to 255-pixel values) into binary image (0 and 1pixel values) by selecting a global threshold that separates the foreground from background. Each pixel is compared with the edge and if it's greater than the edge it's made 1 instead 0. This can be done by using Otsu's method. The process is estimated using the following equation: (t) = (t) (t) +(t) □□□(t) (2) Weights ωi are the possibilities of the 2 classes separated by a threshold t and σi2 variances of those classes.

5.3 MORPHOLOGICAL OPERATION:

The main purpose of this operation is to enlarge and compress the image to the required format. Methods used in this morphological operation are dilation and erosion. While dilation is to enlarge the image, erosion is to compress the image, and this operation aims at displaying a clear view of the image. Morphology may be a broad set of image processing operations that process images supported shapes. Morphological operations apply a structuring element to an input image, creating an output image of an equivalent size. In a morphological operation, the worth of every pixel within the output image is predicated on a comparison of the corresponding pixel within the input image with its neighbors. By choosing the dimensions and shape of the neighborhood, you'll be able to construct a morphological operation that's sensitive to specific shapes within the input image. The basic morphological operations are dilation and erosion. Dilation adds pixels to the boundaries of objects in a picture, while erosion removes pixels on object boundaries. The number of pixels added or off from the objects in a picture depends on the dimensions and shape of the structuring element wont to process the image. In those morphological dilation and erosion operations, the state of any given pixel in the output image is decided by applying a rule to the corresponding pixel and its neighbors in the input image. The rule wont to process the pixels defines the operation as dilation or an erosion. This table lists the principles for both dilation and erosion.

5.4 CONNECTED COMPONENT ANALYSIS:

In binary image, we have to remove the unwanted region based on the area, after that we will count the connected component in the regions. Contiguous regions are called "objects," "connected components," or "blobs." The label matrix containing contiguous regions may appear as if this: 1 1 0 2 2 0 3 3 1 1 0 2 2 0 3 3 Elements of L adequate to 1 belong to the primary contiguous region or connected component; elements of L adequate to 2 belong to the second connected component; then on. Discontinuous regions are regions that might contain multiple connected components. A label matrix containing discontinuous regions might look like this: 1 1 0 1

1 0 2 2 1 1 0 1 1 0 2 2 Elements of L equal to 1 belong to the first region, which is discontinuous and contains two connected components. Elements of L adequate to 2 belong to the second region, which may be a single connected component.

5.5 FEATURE EXTRACTION:

Feature extraction phase extracts the basic components of Tamil characters, such as Height, Width, Horizontal Projection, Vertical Projection, Horizontal Centre, Vertical centre, Horizontal Projection Skewness, vertical Projection Skewness, HCurves, VCurves, number of circles, number of slope lines and branching points. Each feature plays an important role in pattern recognition. So, in order to extract the features from the segmented Tamil character images, first scale the image into common height and width by using bilinear interpolation technique. Hence each image is divided into equal number of horizontal and vertical strips. This linear interpolation technique can do first in x direction and then again in the y direction. Linear interpolation in the x- direction is calculated using the equation

$$F(R_1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \quad (3)$$

Linear interpolation in the *y*-direction is calculated using equation:

$$F(P) \approx \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2). \quad (4)$$

Horizontal centerline: Horizontal centerline is calculated based on scanning the character form left to right of the whole character. Let Horizontal centre is calculated based on the position of the horizontal centreline as follows

$$\text{Hf1} = \sum_{i=2}^{k} |h(i) - h(i-1)| \quad (5)$$

Vertical centre: For exploiting the information coming from the detection of the vertical centreline of the character, generates a new group of the features. To create the second group of the feature vector, the letter image is then scanned from top to bottom with a sliding window. The equation to calculate Vertical centre as follows

$$\text{Vf1} = \sum_{i=2}^{k} |v(i) - v(i-1)| \quad (6)$$

Thus, the Feature extraction describes the relevant shape information contained in a pattern so that the task of classifying the pattern is made easy.

5.6 SEGMENTATION:

After the method of pre-processing, the noise free image is passed to the segmentation phase, where the image is decomposed into individual characters. Algorithm for segmentation: (1) The binarized image is checked for inter line spaces using horizontal and vertical projection technique. (2) If inter line spaces are detected then the image is segmented into sets of portions across the interline gap. (3) The lines within the paragraphs are scanned for horizontal space intersection with reference to the background. Histogram of the image is employed to detect the width of the horizontal lines. Then those suspected lines are scanned vertically for vertical space intersection. Here histograms are wont to detect the width of the words. Finally, the words are decomposed into characters using character width computation.

5.7 MATCHING THE CHARACTER:

The last and final stage of this paper is matching each and every segmented character with the modern Tamil character and checking whether the segmented character is matching with the template character. The matching process can be done by the correlation matching process such that the character can be matched by comparing it pixels with the corresponding image and analyzing the letter and finding the output.

6. CONCLUSION

As mentioned earlier, nature of database and the inventory of tag sets depend on the type of research agenda. This type of multi-layered analysis on large corpora is feasible only with computer-aided methodology. The four cases mentioned in 2 and the few examples given in 4 are a few among 306 hundred of other questions that we have to account for in Tamil. Most of the present POS tag sets don't provide us with a fine-grained annotation scheme. But in computer-aided corpora-based linguistic research, a fine-grained annotation paradigm is essential. In my experimental database on Tamil Inscriptions, I have about 120 tag sets. Three types of rule-based annotations- morph syntactic, syntactic and semantic- are done manually. I have aimed at a fine-grained analysis and so I have opted for a high number of tags. I have also included information like word order types, verbal valence and other minute details that would help to map different changes at morphological, syntactic and semantic levels. I am using, for each sentence the interlinear glossed text (IGT) format, which includes source language text, a morpheme-by-morpheme gloss, and a translation into French or English. Thus, the process of matching with the present Tamil modern letters using the process called 'Corpus Analysis' is our final result such that this process involves all the steps involving the binary conversion, morphological operation, connected component, feature extraction.

7. FUTURE ENHANCEMENT

The on-going research is to illustrate how linguistic corpora can be used as readily available evidence for mapping language development and language variation in time and space. A huge computerized historical corpus would definitely allow a comparative view of the Tamil language at different moments within the history and the exposing of letters in new forms. These data would help us not only to capture different stages of linguistic developments but also will help to check modern theories about variation and alter. The construction of giant electronic corpora in Tamil presents many constraints associated with linguistic theories.

8. REFERENCES

[1] Mingli Zhang and Christian Desrosiers "High-quality Image Restoration Using Low-Rank Patch Regularization a Global Structure Sparsity", IEEE Transactions on image processing, vol:28, no:2, February 2019.

[2] Po-Hsiung Lin, Bo -Hao Chen, Fan-Chieh Cheng, Shih-Chia Huang," A Morphological Mean Filter for Impulse Noise Removal ", Journal of Display Technology, vol:12, no:4, April 2016.

[3] G. Bhuvaneswari, and V. Subbiah Bharathi "An efficient method for digital imaging of ancient stone inscriptions current science, vol: 110, no: 2, Jan. 2016.

[4] Dr.H. S Mohana, Pradeepa R, Pramod N Kammar, Rajithkumar B K, "Identification and Recognition of Ancient Stone In- Scripted Hoysala Characters Using Support Vector Machine (SVM) Model", International Journal of Innovative Research in Technology, vol:1, no:12, Nov 2015.

[5] A.K Chaou, A Mekhaldi and M. Teguar "Elaboration of Novel Image Processing Algorithm for Arcing Discharges Recognition on HV Polluted Insulator Model", IEEE Transactions on Dielectrics and Electrical Insulation vol: 22, no: 2, April 2015.

[6] Dinesh Kumar C, Dr D Surendran M.E Ph.D., "Forgery Image Detection Based on Illumination Colour Classification with Advanced Skin Colour and Edges", International Journal of Scientific & Engineering Research, vol: 5, no: 3, March 2014.

[7] R. AngelinJeniffer, G. Bhuvaneswari "Image Glazing for Thinning of Ancient Tamil Characters"- International Journal of Scientific & Engineering Research, vol: 5, no: 6, June 2014.

[8] Jing guan, Junjinmel, "New class of grayscale morphological filter to enhance infrared building target" IEEE Aerospace and Electronic Systems Magazine , vol: 27 , no: 6 , June 2012.

[9] Ashu Kumar," Line Segmentation Using Contour Tracing", Department of CSE Yadwindra College of Engineering, Talwandi Sabo, Punjab, India, vol: 3, no:1, Jan. 2012.

[10] V.Espinosa-Duro-"Fingerprints thinning algorithm" IEEE Aerospace and Electronic Systems Magazine vol: 18 , no: 9 , Sept. 2003.