

Yoruba Handwritten Character Recognition using Freeman Chain Code and K-Nearest Neighbor Classifier

Jumoke Falilat Ajao^{*1)}, David Olufemi Olawuyi¹⁾, Odetunji O. Odejobi²⁾

¹⁾Department of Computer Science, Kwara State University
Kwara State University Rd, Malete, Nigeria

²⁾Department of Computer Engineering, Obafemi Awolowo University
220005, Ile-Ife, Osun State, Nigeria

How to cite: J. F. Ajao, D. O. Olawuyi, and O. O. Odejobi, "Yoruba Handwritten Character Recognition using Freeman Chain Code and K-Nearest Neighbor Classifier," *Jurnal Teknologi dan Sistem Komputer*, vol. 6, no. 2, pp. 129-134, Oct. 2018. doi: 10.14710/jtsiskom.6.2.2018.129-134, [Online].

Abstract – This work presents a recognition system for Offline Yoruba characters recognition using Freeman chain code and K-Nearest Neighbor (KNN). Most of the Latin word recognition and character recognition have used k-nearest neighbor classifier and other classification algorithms. Research tends to explore the same recognition capability on Yoruba characters recognition. Data were collected from adult indigenous writers and the scanned images were subjected to some level of preprocessing to enhance the quality of the digitized images. Freeman chain code was used to extract the features of THE digitized images and KNN was used to classify the characters based on feature space. The performance of the KNN was compared with other classification algorithms that used Support Vector Machine (SVM) and Bayes classifier for recognition of Yoruba characters. It was observed that the recognition accuracy of the KNN classification algorithm and the Freeman chain code is 87.7%, which outperformed other classifiers used on Yoruba characters.

Keywords – Yoruba characters recognition; Freeman chain code; K-Nearest Neighbor classification

I. INTRODUCTION

Handwritten character recognition has been an extensive area of research in the last three decades [1], [2]. The research in character recognition is popular for its various application potentials in office automation, cheque verification in banks, post offices, and a large variety of business and data entry applications [3], [4]. Other applications involve reading aid for the blind, library automation, language processing, and multimedia design [5]-[8].

Geometrical growth in computational power has enabled the implementation of the current character recognition methodologies [2], [9]-[12]. This also creates an increasing demand on many emerging application domains of character recognition which

Table 1. The Yoruba orthography [23], [24]

High-Tone (Acute sign)	Mid-tone	Low-tone
á é é́ í ó ọ ú	i j	è è̀ ì ò ò̀ ù
Under dot Tier	Eẹ Oọ Sş	
Diagraph Tier	GB gb	
Character Tier	A B D E F G H I J K M N O P R S T U W Y	
	a b d e f g h k l m n o p r s t u w y	

require more advanced methodologies [13]-[17]. The process of optical character recognition (OCR) includes scanning of the text character-by-character, analysis of scanned images, and the translation of the character image into character codes, commonly used in data processing [6], [14].

There are two major problem domains in handwriting recognition. These are online and offline method [10]. The online handwriting recognition involves the automatic conversion of the captured text as it is written on a special digitizer or Personal Digital Assistant (PDA), where a sensor picks up the pen-tip movements as well as pen-up/pen-down switching [11]-[14]. The captured handwritten text is converted into letter codes which are used for the text-processing applications. In the offline technique, the system recognizes the fixed static shape of the character [3], while online character recognition instead recognizes the dynamic motion during handwriting. In offline handwriting recognition, information captured are available on paper, which is digitized through the use of scanner [17]-[21].

Yoruba is a native language spoken by over 40 million people in Nigeria and its diaspora [22]. Yoruba alphabets are tonal, which makes its recognition more difficult than that of printed English alphabet [5], [20], [23]. The Yoruba orthography is represented in Table 1.

Despite, the intensive research on Latin languages, only a few works have been done in the area of indigenous languages [6], [8], [23]. Researchers in handwriting recognition have been considering recognition in different languages as a dissimilar

^{*}) Correspondence author (Jumoke Falilat Ajao)
Email: jumoke.ajao@kwasu.edu.ng

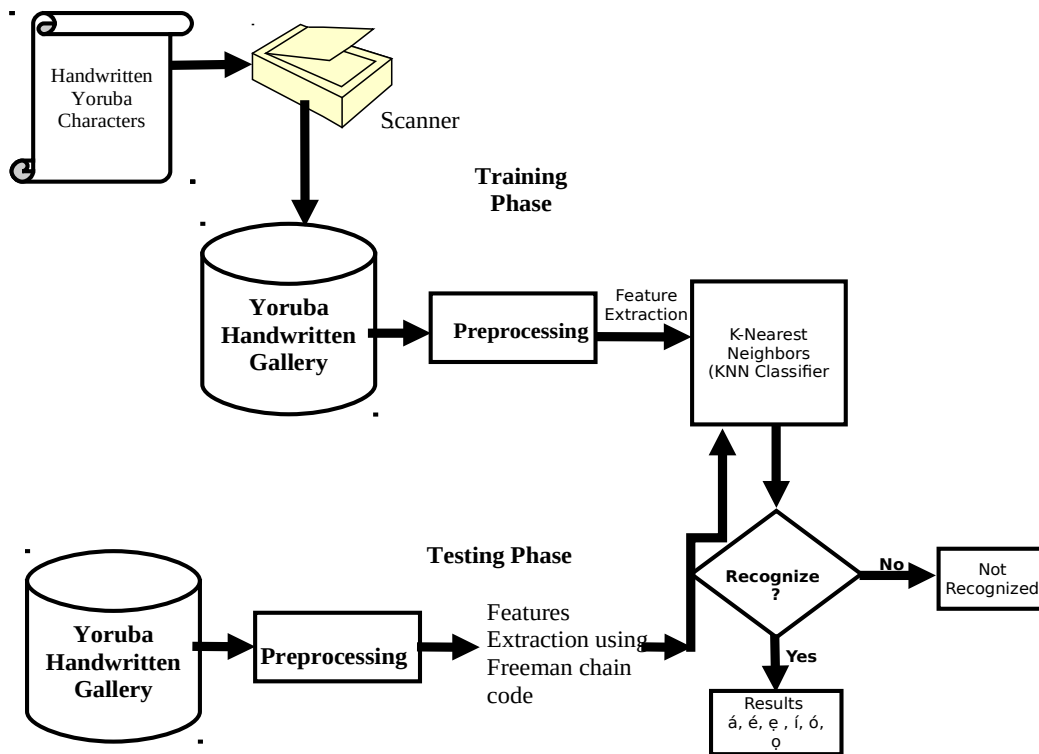


Figure 1. The framework for the Yoruba character recognition

problem. Each language has its own character sets and the language features which make it difficult to develop a general algorithm that works for all the languages.

Many researchers have used Freeman chain code for extracting features of characters or numerals due to its simplicity and the capability for small memory requirement. K-Nearest Neighbors (KNN) was also used because of its simplicity, reduced training time and good performance. Few works that have been done on Yoruba character recognition uses Bayes and SVM. There is a need to test the simplicity of Freeman chain code and the less computational complexity training time of KNN on Yoruba character and to determine its efficiency over other classifiers. This work presents an approach to recognize Yoruba handwritten characters based on Freeman chain code and KNN classifier.

II. METHODOLOGY

The method of the recognition system of Yoruba character includes data acquisition, preprocessing, feature extraction, classification and recognition (Figure 1). The framework of the recognition system began with the collection of handwritten character from various Akomolede indigenous Yoruba writers of a secondary school. The character images were captured using an HP Scanner with 300 dpi resolution. The handwritten characters were scanned in JPEG format, which was converted into Portable Network Graphics (PNG) format and stored in a database using MySQL.

Samples of Yoruba handwritten characters were collected from the “Akomolede” Cultural Group of Lower Niger River Basin Staff College, Ilorin. A guided

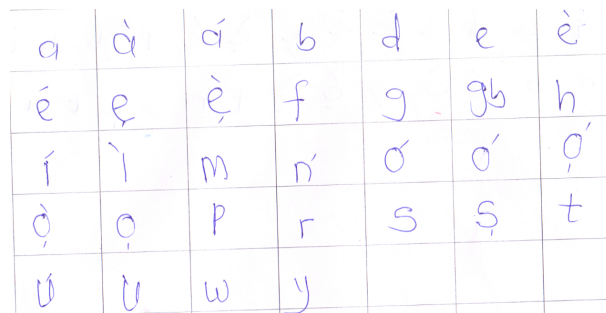


Figure 2. Samples of captured handwritten Yoruba character

template was used to capture various handwriting from the targeted subject, the “Akomolede” Cultural Group are secondary students from various classes (from JSS One to SS Three). They learn and teach other students Yoruba Cultures and Practices of Yoruba in the college. Majority of these students are Yoruba by tribes, a few of them are from other tribes. The handwritten images were scanned using an HP Scanner with 300 dpi resolution (Figure 2). The scanned images were converted into Portable Network Graphics (PNG) format. The scanned handwritten characters were stored in a database.

The series of processes operation was carried out on the captured and digitized image for the enhancement of the digitized image for further processing. In the preprocessing stage, the system carried out data cleaning tasks which include image cropping, noise removal, greyscale conversion, binarization, and

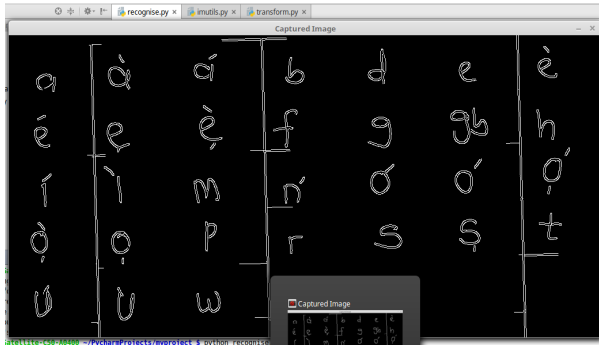


Figure 3. The binarized handwritten image.

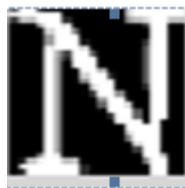


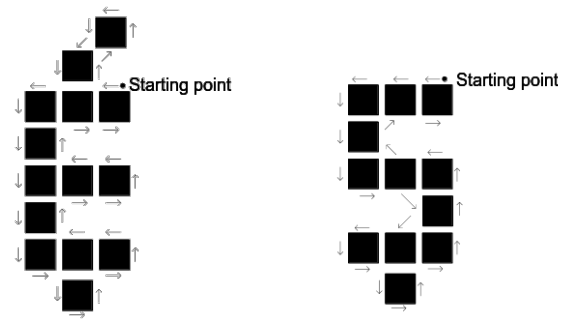
Figure 4. Samples of resized and cropped image

normalization. Features were extracted using the 8-connected Freeman chain code and the KNN was used for the recognition process.

The images were transformed to their discrete form by assigning numbers to the light intensity of each of the points of the discretized images which ranges between 0 and 255. The original RGB image was converted to grayscale and the binarized images were converted to black and white image. Otsu method was used to binarize the image [25]. The binarized image is shown in Figure 3. Digitized image was resized in order to have a minimal input to the KNN to reduce the size of the image to an appreciable size. Cropping removes the unwanted blank space which just an unwanted noise from the four sides of the image in order to have appropriate dimensions. Cropping makes the image to be centrally aligned using the CorelDraw 12 software (Figure 4).

The image was subjected to normalization by converting the various dimensions of images into a fixed dimension to remove handwriting variation and obtain standardized data. The input data was transformed into a reduced representation set of features using a chain code algorithm [26]. The features set extracted the relevant information from the digitized data in order to perform the desired task using this reduced representation instead of the full-size input. The Freeman chain code was used to extract the features of the characters and performed on raw data prior to applying the KNN algorithm on the transformed data in feature space.

In handwritten Yoruba character recognition system, an 8-connected chain code feature extraction techniques was used, which represents each of the characters with 8-connected chain codes. The underdot and the diacritic sign applied to these words make them have different meanings. In this case, it is very difficult to extract



a) E with grave accent and under dot
b) S with under dot

Figure 5. Application of chain code in Yoruba characters

[1	1	1	1	1	0	0	1	0	0	0
0	7	0	7	0	6	7	6	6	5	6
5	5	5	4	5	4	5	4	5	7	7
0	7	0	0	0	0	0	0	0	0	0
7	0	6	6	6	4	4	3	4	4	4
5	4	4	4	3	4	4	3	4	3	4
3	3	3	4	2	3	2	2	2	2	2
2	2	1								

(a) E with grave accent and under dot

[0	0	0	1	1	1	1	1	1	1	1
0	0	6	7	1	2	2	0	0	0	7
7	7	6	6	6	6	5	5	5	4	5
4	5	4	4	4	5	7	7	7	0	6
5	7	0	0	0	0	0	0	0	0	0
0	0	0	1	2	0	6	6	5	4	4
4	4	4	4	4	4	4	4	4	4	4
4	3	4	4	3	4	3	2	3	3	3
4	4	3	4	2	2	2				

(b) S with under dot

Figure 6. The Freeman chain code generated

different objects or components which constitute unique features for each of the individual Yoruba character. The feature extraction techniques extract different profiles such as the character, the starting point, the loop, the curvature, the dot, and the diacritic sign. The extracted features were represented in a feature vector matrix as input to the KNN classifier algorithm.

The application of chain code is illustrated with samples of Yoruba character E with grave accent and under dot and S with under dot (Figure 5). In Yoruba character, the symbol in Figure 5(a) represents vertical line attached to triple linearity with under dot and grave diacritic sign. Figure 5(b) represents dual roundness with dual linearity. Generated Freeman chain codes for the samples of Yoruba characters are shown in Figure 6.

The KNN algorithm was used for the estimation and classification of Yoruba character. According to the KNN estimation technique, Algorithm 1 is performed [14], [16]. x represents the point extracted by the

Table 2. Results analysis of training and testing of Yoruba character

Characters	Each character samples	Training samples	Testing samples	Recognized Characters	Unrecognized Characters	Recognition Rate	Number of Characters
A E É ě Ú Û	61	1	60	57	3	95.00%	6
Á B È ě ě F W	61	1	60	56	4	93.30%	7
N O ó ô	61	1	60	55	5	91.70%	4
À D G Gb Ó	61	1	60	54	6	90.00%	5
H o P	61	1	60	53	7	88.30%	3
Ò ş T U	61	1	60	52	8	86.70%	4
S	61	1	60	51	9	85.00%	1
Y	61	1	60	49	11	81.70%	1
M	61	1	60	47	13	78.30%	1
Í	61	1	60	46	14	76.70%	1
R	61	1	60	45	15	75.00%	1
Ì	61	1	60	44	16	73.30%	1
L	61	1	60	39	21	65.00%	1
I	61	1	60	35	25	58.30%	1

Freeman chain code, d is the Euclidean distance, K is the starting point, and $V(x)$ is the volume of the point that makes up the character, that is $V(x) = \pi r^2$ in the 2D space of the handwritten images. The Euclidean distance is employed and the k -furthest neighbor and x is e . After computing the K -Nearest estimation density function, the KNN classification rule for the recognition of Yoruba character was carried out using Algorithm 2.

Algorithm 1. KNN estimation process

Consider a set of N points, $x_1, x_2, x_3, x_4, \dots, x_n \in Z^1$, that emanates from the Freeman chain code.

- 1: Choose a value k
- 2: Find the distance between x and all training points $x_i, i=1,2,3, \dots, N$, extracted using Freeman chain code, where Euclidean distance is computed as $d = \sqrt{x_{i+1} - x_i^2}$
- 3: Find the k -nearest point to x
- 4: Compute the volume $V(x)$ in which the k -nearest neighbors lie.
- 5: Compute the density estimate by $P(x) = \frac{k}{NV(x)}$

Algorithm 2. KNN classification process of Yoruba character

Given c classes with: $\omega_i, i=1,2, \dots, c$, a point $x \in Z^1$ using Euclidean distance and training point x_i for $i=1 \dots N$ in the 2D space.

- 1: Select among the N training points, Search for k closest neighbor using Euclidean distance measure parameter k , where the starting point that can be choosing from any arbitrary point, which could be even or odd depending on the contour of the character.

- 2: From the K -closest neighbors, identify number k_i of that class that belong to class $\omega_i, \sum_{i=1}^c k_i = k$
- 3: Assign x class $\omega_i, k_i > k_j$, where x is assigned to the class where majority of k -closest neighbor belong.

III. RESULTS AND DISCUSSIONS

In order to validate the developed Yoruba handwritten character recognition system, the handwritten character images were selected from the created Yoruba database, and it was tested on the recognition module. The test data were subjected to preprocessing. Feature extraction of the preprocessed images was carried out and compared with the trained features vector. If the two feature sets are similar, it looks up for the appropriate character in the look-up table and display the correspondence digital equivalent of the handwritten character if found, or brings a False Negative result if not found. The performance of KNN classifier was compared with other results that use different classifiers using different databases. Evaluation metrics was carried out using the recognition rate.

The recognition of Yoruba handwritten character was performed on the Yoruba handwritten database created by the Author. It recognized 37 Yoruba alphabets distinguished based on diacritical marks with 61 captured samples each. Table 2 reveals the testing analysis of all Yoruba alphabets captured for this research work. Each character contains 61 samples which were collected and digitized. Out of the 60 samples collected, one of the samples was used for training the system while the remaining 60 samples were used for testing the system. All samples for training and testing are converted to Portable Network Graphics (.PNG) to enhance the better performance of the system. The results reveal that the highest

Table 3. Accuracy comparison among methods

Authors	Language	Method	Accuracy
Oladele et al. [20]	Yoruba character recognition	SVM	76.70%
Ajao et al. [21]	Yoruba Word Recogniton	HMM	95.60%
Ibraheem and Odejobi [5]	Yoruba Six Uppercase characters recognition	Bayes and decision tree algorithm	94.40%
Wang [14]	English Letters	KNN	80.60%
Our proposed methodology	Yoruba character recognition	KNN	87.70%

recognition rate was 98.3% for the a, e, é, ẹ, ú, and ù alphabets, while the lowest rate was 58.3% for the i alphabet. The recognition rate of these alphabets (i, l, í, and r) are very low. Naturally, the alphabets that seem to be simple and easy to identify are poorly recognized by the system. The recognition rate of the system was evaluated by calculating the mean of the 37 handwritten characters tested. The average cognition rate of the designed system was 87.7%.

The result was compared with other classification algorithms used on Yoruba characters and English letters recognition (Table 3). It was observed that the proposed method outperformed Yoruba character recognition using SVM using 600 handwritten images for Yoruba alphabet [20]. The performance is also better than KNN used in English letter recognition [14] and closed to ones that used combined classifiers of Bayes and decision tree on 6 Yoruba uppercase characters only [5]. This work has achieved the simplicity of KNN with less training time against the complexity involved in multiple classifiers. It is also comparable with [21] that able to recognize Yoruba word, instead of one character, accurately using Hidden Markov Model (HMM).

IV. CONCLUSION

The KNN and Freeman chain code methods for Yoruba character recognition outperformed the one using SVM by achieving average recognition accuracy of 87.7% for 37 Yoruba alphabets distinguished based on diacritical marks. Each alphabet used 61 samples.

REFERENCES

- [1] S. Gunter and H. Bunke. "Ensembles of Classifiers for Handwritten Word Recognition," *International Journal of Document Analysis and Recognition*, vol. 5, no. 4, pp. 224-232, 2003.
- [2] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. San Diego: Academic Press, 2009.
- [3] N. Arica and F. Yarman-Vural, "An Overview of Character Recognition Focused on Off-Line Handwriting," *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 31, no. 2, pp. 216-233, 2001.
- [4] J. F. Ajao, R. G. Jimoh, and S. O. Olabiyisi, "Handwritten Address Destination Recognition using Neural Networks," *Journal of Science, Technology, Mathematics and Education*, vol. 9, no. 1, pp.70-91, 2012.
- [5] A. Ibraheem and O. Odejobi, "A System for the Recognition of Handwritten Yoruba Characters," AGIS Ethiopia, Obafemi Awolowo University, Ile-Ife, Nigeria, 2011. [Online]. Available: <http://www.slideshare.net/aflat/a-system-for-the-recognition-of-handwritten-yoruba-characters>.
- [6] A. A. Desai, "Gujarati Handwritten Numeral Optical Character Reorganization through Neural Network," *Pattern Recognition*, vol. 43, no. 7, pp. 2582-2589, 2010.
- [7] H. Bunke, "Recognition of Cursive Roman Handwriting: Past, Present and Future," in *Proc. of Seventh International Conference on Document Analysis and Recognition*, Edinburgh, 2003, pp. 448-459.
- [8] C. Liu and C. Y. Suen, "A New Benchmark on the Recognition of Handwritten Bangla and Farsi Numeral Characters," *Pattern Recognition*, vol. 42, no. 12, pp. 3287-3295, 2009.
- [9] D. Impedovo and G. Pirlo, "Zoning Methods for Handwritten Character Recognition: A Survey," *Pattern Recognition, Handwriting Recognition, and Other PR Applications*, vol. 47, no. 3, pp. 969-981, 2014.
- [10] M. Cheriet, N. Kharma, C-L. Liu, and C. Y. Suen, *Character Recognition Systems: A Guide for Students and Practitioners*. New York: John Wiley, November 2007.
- [11] O. D. Femwa, "Development of a Writer-Independent Online, Handwritten Character Recognition System Using Modified Hybrid Neural Network Model," PhD. Thesis, Ladoko Akintola University of Technology, Ogbomoso. 2012.
- [12] F. O. Deborah, O. E. Olusayo, and F. O. Alade, "Development of a Feature Extraction Technique for Online Character Recognition System," *Innovative Systems Design and Engineering*, vol. 3, no. 3, pp. 10-23, 2012
- [13] J. O. Adigun, O. D. Femwa, E. O. Omidiora, and S. O. Olabiyisi, "Optimized Features for Genetic Based Neural Network Model for Online Character Recognition," *British Journal of Mathematics & Computer Science*, vol. 14, no. 6, pp.1-13, 2016.
- [14] W. Wang, "Optical Character Recognition, Using K-Nearest Neighbors," arXiv:1411.1442 [cs.CV], Nov. 2014.
- [15] I. S. Shimasaan and E. M. Bukohwo, "An Artificial Network Model for Tiv Character

- Recognition,” *Journal of Emerging trends in Computing and Information Sciences*, vol. 6, no. 10, pp. 573-583, 2015.
- [16] R. Siddhartha and M. Saravanan. “Handwritten Character Recognition using KNN,” *International Journal of Advance Research and Innovative Ideas in Education*, vol. 3, no. 5, pp. 2395-4396, 2017.
- [17] W. L. Wang and M. H. Tang, “A Normalization Process to Standardize Handwriting Data Collected from Multiple Resources for Recognition,” *Procedia Computer Science*, vol. 61, pp. 402-409, 2015.
- [18] N. Azizi, N. Farah, and M. Sellami, “Off-line Handwritten Word Recognition System using Ensemble of Classifier Selection and Features Fusion,” *Journal of Theoretical and Applied Information Technology*, vol. 14, no. 2, pp. 141-150, 2010.
- [19] G. Katiyar, A. Katiyar, and S. Mehruz, “Offline Handwritten Character Recognition System using Support Vector Machine,” *American Journal of Neural Networks and Applications*, vol. 3, no. 2, pp. 22-28, 2017.
- [20] M. Oladele, E. Omidiora, A. Temilola, and A. A. Sobowale, “An Offline Yoruba Handwritten Character Recognition System using Support Vector Machine,” *International Conference of Sciences, Engineering & Environmental Technology (ICONSEET)*, vol. 2, no. 13, pp. 95-103, 2017.
- [21] J. F. Ajao, S. O. Olabiyisi, E. O. Omidiora, and O. O. Okediran, “Hidden Markov Model Approach for Offline Yoruba Handwritten Word Recognition,” *British Journal of Mathematics & Computer Science*, vol. 18, no. 6, pp. 1-20, 2016.
- [22] Omniglot, “The Yoruba Alphabets and Its Pronunciation,” [Online]. Available: <http://www.omniglot.com/writing/yoruba.htm>
- [23] J. F. Ajao, S. O. Olabiyisi, O. O. Elijah, and O. O. Odetunji, “Yoruba Handwriting Word Recognition Quality Evaluation of Preprocessing Attributes using Information Theory Approach,” *International Journal of Applied Information Systems*, vol. 9, no. 1, pp. 18-23, 2015.
- [24] A. Bamgbose, *Yoruba Orthography*. Ibadan University Press, 1976.
- [25] H. J. Vala and A. Baxi, "A Review on Otsu Image Segmentation Algorithm," *International Journal of Advanced Research in Computer Engineering and Technology (IJARCET)*, vol. 2, no. 2, pp. 96-106, 2013.
- [26] E. Bribiesca, “A New Chain Code,” *Pattern Recognition*, vol. 32, no. 2, pp. 235-251, 1999.