ФИЗИКО-МАТЕМАТИЧЕСКИЕ И ТЕХНИЧЕСКИЕ НАУКИ

УДК 004.3
МРНТИ 28.23.15; 28.23.29; 28.23.24; 28.23.33

COMPARATIVE ANALYSIS OF RECOGNITION ALGORITHMS FOR HAND GESTURES ON THE BASIS OF VARIOUS REPRESENTATIONS OF IMAGES

Y. AMIRGALIYEV[1,2], A. AITIMOV[1,2], B. AMIRGALIYEV[3], B. KYNABAY[2]

[1]Suleyman Demirel University
[2]Institute of Information and Computational Technologies
[3]International University of Information Technologies

Abstract: *At this time, the world has created many different algorithms for recognizing hand gestures. In this paper, the authors reviewed and proposed various gesture recognition algorithms to determine the best among them in terms of speed and quality of recognition. The algorithms like K Nearest Neighbors, Decision Trees, Logistic Regression was compared with two methods of representation pictures. The results showed us that the Logistic Regression with using Raw pixel method better than other algorithms.*

Keywords: *hand gesture, sign gesture, KNN, D-Tree, Logit, machine learning, ML, image recognition and processing*

СРАВНИТЕЛЬНЫЙ АНАЛИЗ АЛГОРИТМОВ РАСПОЗНАВАНИЯ ЖЕСТОВ РУК НА БАЗЕ РАЗЛИЧНОГО ПРЕДСТАВЛЕНИЯ ИЗОБРАЖЕНИЙ

Аннотация: *В данное время в мире создано множество различных алгоритмов для распознавания жестов рук. В данной работе авторами рассмотрены и предложены различные алгоритмы распознавания жестов, чтобы определить наилучшее среди них с точки зрения скорости и качества распознавания. Алгоритмы, такие как K Nearest Neighbors, Decision Trees, Logistic Regression, сравнивались с двумя методами представления изображений. По результатам исследовательской работы Logistic Regression с использованием необработанного пиксельного метода показывает наилучший результат по скорости и по качеству распознавания по сравнению с другими алгоритмами.*

Ключевые слова: *жест рук, определение жестов, KNN, D-Tree, Logit, машинное обучение, распознавание и обработка изображений*

КЕСКІНДЕРДІҢ ӘРТҮРЛІ БЕРІЛІСІНДЕ ҚОЛ ҚИМЫЛЫН ТАНУДЫҢ АЛГОРИТМДЕРІН САЛЫСТЫРМАЛЫ ТАЛДАУ

Аңдатпа: *Қазіргі таңда қол қимылдарын анықтауға арналған әртүрлі көптеген алгоритмдер ойлап табылған. Бұл ғылыми жұмыста жылдамдық пен қол қимылдарын тану сапасы жағынан ең тиімді алгоритм анықталатын болады. K Nearest Neighbors, Decision Trees, Logistic Regression сияқты алгоритмдер ұсынылған кескіндердің екі әдіспен берілуін қолдана отырып өзара салыстырылды. Зерттеу жұмысының қорытындысы бойынша Raw pixel әдісін пайдалану арқылы Logistic Regression алгоритмі жылдамдық және сапа жағынан оң нәтиже көрсетті.*

Түйінді сөздер: *қол қимылдыры, қимылдарды тану, KNN, D-Tree, Logit, машиналық оқыту, кескінді анықтау және өңдеу*

Introduction

Communication between people is an integral part of our life. People use it unnoticed by it over-importance. It's even difficult to imagine what would happen to a social society if people couldn't be able to convey their thoughts, show emotions, express what they want. Because lack of communication leads to very bad consequences [1]. The problem with a lack of communication is often found in people with disabilities namely in the deaf-and-dumb, and which lead to loneliness and isolation from the social society [2]. According to the World Health Organization - WHO (WHO, 2017), 5% of the world's population have hearing loss that can be mild, moderate, severe or profound. This means that approximately 360 million people struggle every day to communicate [3].

Most people who have difficulty hearing, use the language of gesture for the concept of thoughts or actions of the interlocutor and to convey their opinions. However, at the moment in connection with globalization, the latest technologies are gradually switching to automation and control of objects with the help of vocal cords[4]. And due to the fact that people with disabilities find it difficult to use advanced technologies, scientists are introducing the recognition of hand gestures. From this situation, several questions emanate, such as which algorithm is best suited for solving this problem and which method is best to use?

In this scientific work, we determined which of the machine learning algorithms (KNN, D-Tree, Logit) would be best to use and what methodology should be used for the qualitative and simultaneous rapid recognition of hand gestures with the help of several experiments. As data used dataset "Hand Reader Dataset" that is in open source[5]. The result of our research work is that for a small amount of data the most effective algorithm is logistic regression using the raw pixel method.

Section, description of used dataset, describes the purpose of the data used to compare algorithms. Description of the algorithms and methods used for the experiment is provided in the section of review of algorithms and methods. Experiment and results section is about the experiments and their results. In the last section we summarized our paper and wrote about future work.

Description of used dataset

Models simulations performed over machine learning (ML) framework for number of different algorithms over the dataset of Ghassem Tofighi that named "Hand Reader Dataset" in order to find out the best possible way of formatting pictures for training the models and to reveal the true potential of these algorithms. This section provides an information about the key details of the dataset of pictures and possible ways of their application in the research work.

Actual dataset consists of 500 images that contains hand postures of 10 different types: 10 letters of American Sign Language alphabet [6]. However, only 4 of the hand postures (A, B, C, D) were used in current research paper. There are 6 differently sized collections for each letter, created manually, each contains 50 images. Format of the input data is JPG. Each sign is shown through number of images that are different in terms of scale and form of hand posture and each of them are unique. Background is black for every picture, but can differ in light condition.

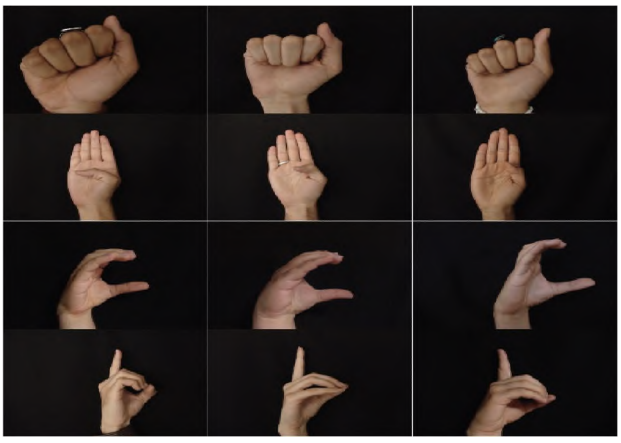

Fig.1. Hand Reader Dataset (A,B,C and D respectively)

Review of algorithms and methods

A. Algorithms

a. K Nearest Neighbors (KNN) algorithm is one of the first simple supervised learning machine learning algorithms[7]. The logic behind this method is to find a predefined number of training samples closest in distance to the new point, and predict the label from these given data-points. Despite its simplicity, nearest neighbors has been successful in a large number of classification and regression problems[8]. As a distance metric generally the Euclidean distance measure is used.

b. Decision Trees (D-Tree) is a supervised learning method that is used for classification and regression[9]. The feature is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features[10]. This method has some advantages like being simple to understand and easy to interpret and also trees can be visualized and requires little data preparation. The method is based on information theory paradigm[11].

c. Logistic regression (Logit) is a part of regression models where the output value is binary or dichotomous. The prediction curve is S-shaped and based on a sigmoid function [12]. Because of non-linear nature this algorithm shows one of the best results on getting the classification model for the data[13].

B. Methods

In gesture recognition, performance of results depends on given input data and classification algorithm[14] and this section take attention to first one. To get good results it needs good input data or data must be prepared to algorithms. In computer science, its called image preprocessing[15]. The image preprocessing used to avoid noises, unneeded part of data using algorithms[16]. This work describes 2 image preprocessing methods Raw pixels and Histogram. Both methods crops important part of dataset images and resize it to get more accurate and faster results. As dataset images has black background there is no need to remove noises.

a. Raw pixels method is one of the simplest method[17] as it takes images as matrix of pixels. The matrix converted to feature vector and append to matrix as one column. All images stored in matrix where one column is one image and one pixels has red, green, blue attributes.

b. The next method Histogram returns similar matrix, but before it converts image to HSV format. Then using openCV algorithm calculates histogram values. Taken results normalized to get better result.

Experiment and Results

This section evaluates and compares performance and execution time of three machine learning algorithms including KNN, Logit and D-Tree using Raw pixel and histogram mentioned in section III. Experiments were done in python platform. Before simulation all dataset images were cropped to get hand gestures and rescaled into six group of different sizes with 20x15, 24x18, 28x21, 32x24, 36x27 and 40x30.

Table 1. Raw pixel method, accuracy.

Raw pixels	KNN		Logit		D-Tree	
SIZE	RR	TIME	RR	TIME	RR	TIME
20x15	95	0,107	97	0,076	83	0,043
24x18	97	0,109	97	0,117	78	0,053
28x21	97	0,11	98	0,098	80	0,053
32x24	97	0,106	98	0,086	84	0,047
36x27	96	0,106	98	0,122	86	0,053
40x30	96	0,113	98	0,3	78	0,158

Table 1 shows accuracy and time of the algorithms using Raw pixel. According to the table KNN shows good accuracy between 95%-97% in different sizes. It is a stable algorithm considering that in six sizes accuracy difference is just 2%. Also, KNN has approximately constant time 0.1 second. Comparing with KNN, Logit has more accuracy and its recognition rate is between 97% and 98%. This means that Logit is also stable. In addition 98% accuracy is the best result in this experiment. However, it is the slowest algorithm as average time is 0.13 second comparing with KNN 0.109 seconds and D-Tree 0.06 seconds(see Fig 3). D-Tree is two

times faster than Logit and the fastest algorithm, but less accurate as its highest result is 86%. Concluding overall experiment using method A bar chart(see Fig 2) shows average accuracy of algorithms KNN, Logit and D-Tree are 96.3%, 97.67 and respectively 81.5%

Table 2. Histogram method, accuracy.

Histogram	KNN		Logit		D-Tree	
SIZE	RR	TIME	RR	TIME	RR	TIME
20x15	62	0,103	52	0,001	59	0,001
24x18	62	0,103	52	0,001	59	0,001
28x21	62	0,105	52	0,003	59	0,001
32x24	62	0,108	52	0,003	59	0,002
36x27	62	0,108	52	0,004	59	0,003
40x30	62	0,108	52	0,004	59	0,002

In another experiment using histogram KNN shows a little improvement in time 0.105 seconds in average as given in table 2. But, Logit and D-Tree demonstrates high speed with time 0.0016 seconds and 0.0026 seconds respectively. In this method Logit 50 and D-Tree 37.5 times works faster than first experiment(see Fig 3). In addition all three algorithms shows constant accuracy in different sizes(see Fig 2). This huge progress in time and stability in accuracy in method B makes it unique, but low accuracy makes it unusable. All three algorithms failed this experiment as recognition rate for KNN 62%, Logit 52% and D-Tree 59%.

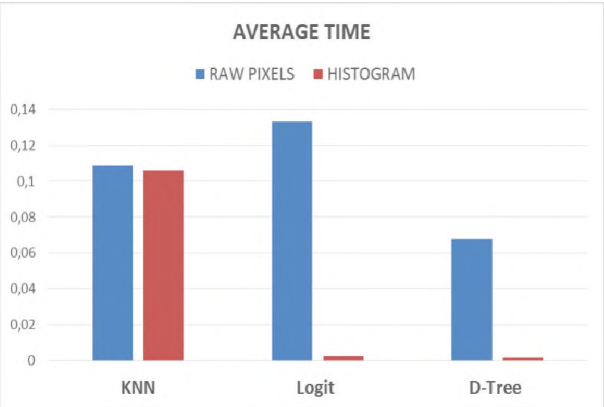

Fig.2. Average accuracy

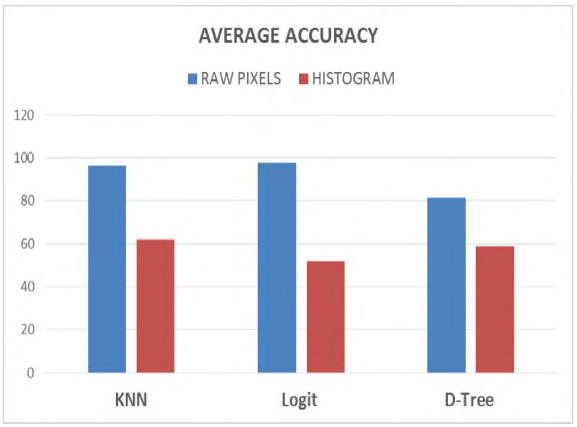

Fig.3. Average time

Conclusion

Gesture plays one of the important role in communication in social life. Recent researches show that for more than 300 million people gesture is only way of communication. In addition, latest technologies are switching to speech controlling system, which makes impossible to use for deaf people[18]. In this work, we tried concentrate to this problem by comparing KNN, Logit and D-Tree machine learning algorithms using Raw pixel and histogram. As dataset used 4 hand postures with 6 different sizes and 50 images each with black background. Main goal is define which algorithm and method has the best performance both in recognition rate and time. Experiment result show that accuracy of algorithms are better when using Raw pixel and Logit with 98%, however needs more time. Comparing with Raw pixel, in histogram all three algorithms has good time and stable results, but maximum accuracy is 62%. So, it is recommended to use Logit in small datasets. In future work needs update Raw pixel method to improve its speed in big datasets.

This work is supported by grant from the Ministry of Education and Science of the Republic of Kazakhstan within the framework of the Project «AP05132648 Creation of verbal-interactive robots based on modern speech and mobile technologies».

REFERENCES

1. Davi Hirafuji Neiva, Cleber Zanchettin. "Gesture recognition: A review focusing on sign language in a mobile context", Expert Systems with Applications, 2018
2. Shahriar Shamiluulu, Moussa Mahamat Boukar, Zulfiya Yussupova. "Medical tool for assisting patients in Kazakhstan polyclinics", 2017 13th International Conference on Electronics, Computer and Computation (ICECCO), 2017
3. M. Stampar, K. Fertalj. "Artificial intelligence in network intrusion detection", 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2015
4. E. M. Simonsick, J. M. Guralnik, S. Volpato, J. Balfour, L. P. Fried, Just get out the door! importance of walking outside the home for maintaining mobility: findings from the women's health and aging study, Journal of the American Geriatrics Society 53 (2) (2005) 198–203.
5. S. E. Hardy, Y. Kang, S. A. Studenski, H. B. Degenholtz, Ability to walk 1/4 mile predicts subsequent disability, mortality, and health care costs, Journal of general internal medicine 26 (2) (2011) 130–135.
6. Global recommendations on physical activity for health. world health organization. URL http://www.who.int/dietphysicalactivity/factsheet recommendations/en
7. Physical activity guidelines for americans. u.s. department of health and human services. URL http://health.gov/paguidelines
8. Nike+ run club app. URL https://www.nike.com/us/en us/c/nike-plus/running-app-gps
9. Runkeeper app. URL https://runkeeper.com/
10. Mapmyrun app. URL http://www.mapmyrun.com/
11. T. Park, J. Lee, I. Hwang, C. Yoo, L. Nachman, J. Song, E-gesture: a collaborative architecture for energy-efficient gesture recognition with hand-worn sensor and mobile devices, in: Proceedings of the ACM SenSys, ACM, 2011, pp. 260–273.
12. J. Viterbi, Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, in: The Foundations Of The Digital Wireless World: Selected Works of AJ Viterbi, World Scientific, 2010, pp. 41–50.
13. K. Murao, T. Terada, A recognition method for combined activities with accelerometers, in: Proceedings of the ACM UbiComp, ACM, 2014, pp. 787–796.
14. H. Junker, O. Amft, P. Lukowicz, G. Troster, Gesture spotting with body-worn inertial sensors to detect user activities, Pattern Recognition ¨ 41 (6) (2008) 2010–2024.
15. Parate, M.-C. Chiu, C. Chadowitz, D. Ganesan, E. Kalogerakis, Risq: Recognizing smoking gestures with inertial sensors on a wristband, in: Proceedings of the ACM MobiSys, ACM, 2014, pp. 149–161.
16. H. Zhao, S. Wang, G. Zhou, D. Zhang, Ultigesture: A wristband-based platform for continuous gesture control in healthcare, Smart Health.
17. P. Alfeld, A trivariate clough—tocher scheme for tetrahedral data, Computer Aided Geometric Design 1 (2) (1984) 169–181.
18. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, The weka data mining software: an update, Proceedings of the ACM SIGKDD 11 (1) (2009) 10–18.