Handwritten recognition of Hiragana and Katakana characters based on template matching algorithm

Y Septiana*, A Mu\textsuperscript{li}ani, D Kurniadi and H Hasanudin

Department of Informatics, Sekolah Tinggi Teknologi Garut, Jalan Mayor Syamsu 1, Garut 44151, Indonesia

*ys	extsuperscript{e}ptiana@sttgarut.ac.id

Abstract. Japanese has become one of the most popular foreign languages in Indonesia. From a survey conducted by The Japan Foundation in 2017 shows that Indonesia is the first largest student in Southeast Asia, therefore those interested in learning Japanese are quite popular in Indonesia. One of the things learned in Japanese is about writing Hiragana and Katakana characters, each of which has 46 standard characters. The research aims to implement a template matching algorithm for handwritten recognition of Hiragana and Katakana characters, which can later be implemented in application design for learning Hiragana and Katakana character writing. Through the application of template matching algorithm, the accuracy level of handwritten recognitions pattern matching reaches 89.8%, so the algorithm will be very suitable if implemented for Hiragana and Katakana characters writing learning applications.

1. Introduction

Based on a survey conducted by The Japan Foundation in 2017 shows that Indonesia is the first largest student in Southeast Asia and the second largest in the world with a total of 745,125 people, who have an interest in learning Japanese [1]. This shows the interest of the Indonesian people towards learning Japanese is quite large. Japanese is one language that is not easy to learn, a common thing that becomes difficult in learning Japanese is the language rules it learns. In learning Japanese there is one scope that is considered the most difficult, namely learning to write characters in Japanese. There are several types of character letters in Japanese, including Hiragana (characters used to write original Japanese vocabulary) and Katakana (characters used to write absorption vocabulary from foreign languages) [2–4].

Template matching is one of the algorithms that are widely used to recognize patterns because its application is very simple and easy. The input image will be compared with the image contained in the database, then the image is compared and calculated how many points best fit the template [5–8]. The level of suitability of the input image with the image in the database can be calculated with the smallest error value [9–12]. The template with the fewest error values is the template that best matches the entered image being compared. The template matching algorithm can be used for handwritten character recognition with a high degree of accuracy [13–15].

This research aims to apply a template matching algorithm for handwritten recognition of Hiragana and Katakana characters, which can later be implemented to the design of educational applications for learning to write Hiragana and Katakana characters based on Android, so that it can provide learning through writing and character checking features.
2. Methods
Handwritten character recognition uses a template matching algorithm, which is a simple method and is widely used to recognize patterns based on shape, size and orientation. Template matching algorithm is often used for recognizing text patterns, symbols and images [5,6,14]. The steps to use template matching algorithm are as follows:

- Change the input image to binary image to facilitate the letter recognition process;
- Separate characters in the image;
- Adjust the resolution of the input image with the image in the database;
- Perform the character recognition process with the Template Matching equation.

The framework of this research can be seen in Figure 1.

![Research framework](image)

Based on the research framework in Figure 1, the stages of implementing the Template Matching Algorithm into the Hiragana and Katakana character writing learning application Android-based are as follows:

- Creation of character patterns;
- Making template matching algorithm script;
- Character recognition check;
- Develop the android-based application.

3. Results and discussion
3.1. Creation of character patterns
The stages of character pattern making are done by creating an area in the Two-Dimensional Array with Order 7x7 according to the predetermined characters [5], as shown in Figure 2.
Figure 2. Character pattern.

From the character patterns that are formed based on Figure 2, then converted into binary images to facilitate the recognition process. Each area in the array that is not used will be given the value 0 in red area, while the area used will be given value 1 in green area.

3.2. Making template matching algorithm script

The script for the character recognition process, as shown in Figure 3.

```
1 arrE, arrT : array of integer;
2 e, count : integer;
3 arrE = array Entry;
4 arrT = array in Database;
5 e = 0;
6 count = 0;
7 For i = 0; i < 7; i++ do
8 For j = 0; j < 7; j++ do
9     count = Math.abs(arrE [i, j] - arrT [i, j]);
10 End For;
11 End For;
12 If e <= 5 then
13     Output("Pattern : Match");
14 Else
15     Output("Pattern : Not Match");
16 End if;
```

Figure 3. Template matching algorithm script.

Based on the Template Matching Algorithm Script in Figure 3, the character recognition process occurs with the following steps [6,14]:

- Create an empty array to accommodate the character patterns created by users and those in the database;
- Create a variable to hold pixel counts in the array and count the number of errors;
- Fill in the empty array arrE with an array of patterns entered from the user;
- Fill in the empty array arrT with an array of patterns residing in the database according to user choice;
- Fill variable to accommodate errors with 0;
- Fill variable to count each pixel in the array with 0;
- Perform the process of repeating as many pixels as there are in arrE, then fill in the calculated variable with the absolute value of the reduction in the value of each pixel in the arrE with the value of each pixel in the arrT;
- Fill in the variable e with the value of the variable e itself then add up the calculated variable;
- If the error value is less than or equal to 5 then the pattern is deemed appropriate, but if the error value is more than 5 then the pattern is considered to be inappropriate.

3.3. **Character recognition check**

At this stage, it is done by checking the Template Matching algorithm where the pattern created will be matched with the existing patterns in the database whether it is suitable or not. If the error is less than 10.2%, then the pattern is considered appropriate, but if the error is more than 10.2%, then the pattern is considered inappropriate. The Character Recognition Check, as shown in Figure 4.

![Figure 4. Character recognition check.](image)

The illustration of the character recognition check based on algorithm on Figure 4 as follows:

3.3.1. **Handwritten patterns of Hiragana and Katakana characters from users.**

|     | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
|-----|---|---|---|---|---|---|---|
|     | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
|     | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
|     | 1 | 1 | 1 | 0 | 0 | 1 | 0 |
|     | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
|     | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
|     | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

**Figure 5.** Pattern of Hiragana and Katakana characters from users.
3.3.2. Hiragana and Katakana character patterns stored in the database.

|   |   |   |   |   |   |
|---|---|---|---|---|---|
|   | 1 | 1 | 1 | 1 | 1 |
| 1 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 0 | 0 | 1 |
| 1 | 1 | 0 | 0 | 1 | 1 |
| 1 | 0 | 0 | 0 | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 | 1 |

Figure 6. Hiragana and Katakana character patterns stored in the database.

3.3.3. Check the suitability of Hiragana and Katakana character patterns using the following equation [5].

\[
\text{Min } e = \sum (I_{x,y} - T_{x,y})^2
\]  

(1)

Figure 7. Check the suitability of Hiragana and Katakana character patterns.

Based on these illustrations, the result of checking for pattern errors is the value of \( e = 2 \) (the percentage of errors is 4.08%). The results of the pattern matching are obtained from the calculation of the number of matching and non-matching patterns by comparing the area in the Two-Dimensional Array between the character patterns from users in Figure 5, with the character patterns stored in the database in Figure 6, then checking the suitability of the character patterns with using the template matching equation as illustrated in Figure 7. Because the percentage of errors is less than 10.2%, the pattern is match (suitable).
3.4. Develop the Android-based application

![Android-based application](image.png)

Figure 8. Android-based application.

The result of this research is an Android application for learning Japanese by applying the Template Matching algorithm, which aims to provide learning for people who want to learn the characterization of Japanese characters including Hiragana and Katakana.

4. Conclusions

Japanese learning application by applying the Algorithm Matching Template can help students and the public to know the basic language of Japanese characters because this application is equipped with the following features, namely the writing and checking features designed using the Algorithm Matching Template, by using this algorithm checking on the feature handwritten recognition for Hiragana and Katakana characters have a higher level of accuracy compared to other algorithms, which is 89.8%.

Acknowledgment

Authors wishing to acknowledge Sekolah Tinggi Teknologi Garut that supports and funds this research publication.

References

[1] The Japan Foundation 2017 Survey Report on Japanese-Language Education Abroad
[2] Das S and Banerjee S 2012 A pattern recognition approach for handwritten hiragana character identification IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2012
[3] Nam N T and Hung P D 2019 Padding methods in convolutional sequence model: An application in Japanese handwriting recognition ACM International Conference Proceeding Series (New York, New York, USA: Association for Computing Machinery) pp 138–42
[4] Das S and Banerjee S 2015 An Algorithm for Japanese Character Recognition I.J. Image, Graph. Signal Process. 1 9–15
[5] Brunelli R 2009 Template matching techniques in computer vision: theory and practice (John Wiley & Sons)
[6] Hossain M A and Afrin S 2019 Optical Character Recognition based on Template Matching Glob. J. Comput. Sci. Technol. 19 30–5
[7] Ahmed M and Abidi A I 2019 Performance comparison of ANN and template matching on English character recognition Int. J. Adv. Res. Ideas Innov. Technol. 5 367–72
[8] Baswardono W, Kurniadi D, Mulyani A and Arifin D M 2019 Comparative analysis of decision tree algorithms: Random forest and C4.5 for airlines customer satisfaction classification J. Phys. Conf. Ser. 1402
[9] Pragathi M A, Priyadarshini K, Saveetha S, Banu A S and Mohammed Aarif K O 2019
Handwritten Tamil Character Recognition Using Deep Learning Proceedings - International Conference on Vision Towards Emerging Trends in Communication and Networking, ViTECoN 2019 (Institute of Electrical and Electronics Engineers Inc.)

[10] Wang M, Li Z, Si F and Guan L 2019 An improved image thinning algorithm and its application in Chinese character image refining Proceedings of 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2019 (Institute of Electrical and Electronics Engineers Inc.) 1870–4

[11] Tiwari U, Jain M and Mehfuz S 2019 Handwritten character recognition—an analysis Lecture Notes in Electrical Engineering vol 509 (Springer Verlag) 207–12

[12] Ahmed S Bin, Naz S, Swati S and Razzak M I 2019 Handwritten Urdu character recognition using one-dimensional BLSTM classifier Neural Comput. Appl. 31 1143–51

[13] Bong C W, Xian P Y and Thomas J 2020 Face Recognition and Detection Using Haars Features with Template Matching Algorithm Advances in Intelligent Systems and Computing vol 1072 (Springer) 457–68

[14] Kalina D and Golovanov R 2019 Application of template matching for optical character recognition Proceedings of the 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2019 (Institute of Electrical and Electronics Engineers Inc.) 2213–7

[15] Sharma A K, Thakkar P, Adhyaru D M and Zaveri T H 2019 Handwritten Gujarati Character Recognition Using Structural Decomposition Technique Pattern Recognit. Image Anal. 29 325–38