

Principal Component Analysis Method for Image Segmentation of Smart Meter Digital

Husnawati^{1,*} Shinta Puspasari¹ Rian Rahmanda Putra²

¹ Department of Computer Science, Indo Global Mandiri University, Palembang, Indonesia

² Department of Computer Science, Politeknik Negeri Sriwijaya, Palembang, Indonesia

*Corresponding author. Email: shinta@uigm.ac.id

ABSTRACT

The digital kWh meter is a device that can measure the consumption of electrical power in electrical equipment. This paper presents an approach to recognize numeric characters on the kWh meter image by implementing principal component analysis for segmentation. Research methodology contains two phases, pre-processing and recognition for image processing. The performance of the proposed approach is evaluated by developing the prototype of a nonreal-time system to recognize numeric characters contained on image and use a dataset of Smart Meter Indonesia Model of kWh Meter images as input for testing. In terms of effectiveness, the experimental result shows accuracy up to 90% on average for kWh Meter digital recognition. It represents correctness in reading the value measured by kWh Meter.

Keywords: *Smart Meter Digital, Image Processing, PCA*

1. INTRODUCTION

The Electricity becomes the needs of humans. Nowadays it is used as energy to support daily activities such as lighting, empower electric devices, and many others. It can be found in every place, especially at home and building. An electric meter is needed to measure the amount of electrical energy supplied on it. The common one is the kWh Meter. It records the value measured by its meter and the invoice for the electricity usage generated based on this. Traditional electric meters are electro mechanical devices [1], equipped with a series of mechanical dial faces, read by a man, and might present human errors when reading the data. The efficiency and effectiveness of the electric meters reading method need improvement in modern life. A modern dials area electronic devices and can transmit its readings automatically.

The rapid development of technology enabling automation in reading kWh Meter for reducing manual efforts and minimizing errors. This is part of the smart grid in the electronic power industry. The smart grid means the power providers and consumers optimize energy efficiency and real-time information exchange in both directions, by integrating information and communication technology into the power grid [2]. Anywhere and anytime electric usage can be controlled by the consumer.

In Indonesia, the electricity retailed by the government company namely *Perusahaan Listrik Negara* (PLN). People pay to PLN for the bill of electric usage based on the kWh Meter measurement. A man meter read the kilowatt meter then record and send it to the server using a mobile device. Human error might be presented when reading and recording data. It gives disadvantages to both PLN and consumer or end-user such as deficiency and over payment [3]. PLN has a commitment to deliver excellent service to the end-user. The smart grid seems to become the future plan for PLN to improve efficiency, reliability, economics, and sustainability of production and distribution of electricity [2] in Indonesia.

An approach to read numeric characters on kWh Meter automatically is needed to overcome the existing problem due to human errors and improvement inefficiency.

Numeric characters recognition concerned to 10 numeric digits 0-9. Many representation approaches have been proposed in the previous work. Such as Support Vector Machine [2][4], K-Nearest Neighbors algorithm [1][5], Neural Network [3][6], and Hausdorff Distance [7]. The result has shown its effectiveness in reading the character on the image with high accuracy. This study proposes a well-known technique, Principal Component Analysis, for determining an optimal basis

for data reconstruction [8] and its effectiveness has been proved in many cases [9][10][11] not only on the 2D image but also video. The performance of principal component analysis for numeric character segmentation on kWh Meter images will be evaluated using a prototype which is design for the offline system, non real-time system, to recognize numeric characters contained in the image. In terms of effectiveness, the prototype is tested using a dataset contain 640 x480 pixels images of the Smart Meter Indonesia Model of kWh Meter. The input images in specific size and format are processed using a common image preprocessing method and recognized the numeric character on it as the output of the system and represent the usage of electricity measured by kWh Meter. the image used in this study is taken during day time from one angle shooting and does not cover a variety angle that might exist when the image is taken. The rest of the paper is organized as follows: Section 2 gives information on the principal component analysis. Section 3 discusses the detail steps of recognition methodology proposed in this study by performing PCA. Section 4 describes the experimental result of and finally concluded at the end of the paper.

2. LITERATUR REVIEW

The principal component analysis is a well-known technique for determining an optimal basis for the reconstruction of data [8]. Method of Principal Component Analysis based on Euclidean space has been widely used for the character recognition method. The basic principle of the Decision of eigenspace is using a principal component of the data (images) and maintaining the characteristics of the data set, namely eigenvalues and eigenvectors. These features have been proven its effectiveness for classification, face image classification for example, namely eigenface method [12,13].

Principal component analysis (PCA) has been called one of the most valuable results from applied linear algebra. PCA is used abundantly in all forms of analysis - from neuroscience to computer graphics- because it is a simple, non-parametric method of extracting relevant information from confusing data sets. With minimal additional effort PCA provides a road map for how to reduce a complex data set to a lower dimension to reveal the sometimes hidden, simplified dynamics that often underlie it [14]. An algebraic solution to PCA can be derived using linear algebra based on an important property of eigenvector decomposition. Once again, the data set is X , a $m \times n$ matrix, where m is the number of measurement types and n is the number of data trials. The goal is summarized as follows. Find some orthonormal matrix

P where $Y=PX$ such that $SY = \frac{1}{n-1}YY^T$ is

diagonalized. The row of P is the principal components of X . It begins by rewriting Sy in terms of the variable of choice P .

$$\begin{aligned} Sy &= \frac{1}{n-1}YY^T = \frac{1}{n-1}(PX)(PX)^T = \frac{1}{n-1}(PX)X^T P^T \\ Sy &= \frac{1}{n-1}(PXX^T)P^T = \frac{1}{n-1}PAP^T \end{aligned} \quad (1)$$

$A = XX^T$, where A is symmetric. For a symmetric matrix A , it provides:

$$A = EDE^T \quad (2)$$

where D is a diagonal matrix and E is a matrix of eigenvectors of A arranged as columns. The matrix A has $r = m$ orthonormal eigenvectors where r is the rank of the matrix. The rank of A is less than m when A is degenerate or all data occupy a subspace of dimension $r = m$. Maintaining the constraint of orthogonality by selecting $(m - r)$ additional orthonormal vectors to "fill up" the matrix E . These additional vectors do not effect the final solution because the variances associated with these directions are zero. Selecting the matrix P to be a matrix where each row P_i is an eigenvector of XX^T . By this selection, $P = EY^T$. Substituting into Eq. 2, we find $A = P^T DP$. Then Sy can be evaluated as follows by eq.3.

$$Sy = \frac{1}{n-1}D \quad (3)$$

It is evident that the choice of P diagonalized Sy . This is the goal of the PCA method. It can summarize the result of PCA in the matrices P and Sy . The principal components of X are the eigenvectors of XX^T : or the rows of P . The i^{th} diagonal value of SY is the variance of X along P_i . In practice computing, PCA of a data set X entails (1) subtracting off the mean of each measurement type and (2) computing the eigenvectors of XX^T .

3. RESEARCH METHODOLOGY

At the beginning of this study, a dataset contains a Smart Meter Model of kilowatt-hour images that will be used for testing the recognition effectiveness is initialized. These images have a specific format in JPEG, 640 x 480 pixels size, and specific features (its

angle shooting and focus on the value measured by the kWh Meter) as illustrated in Figure. 1.



Figure 1. kWh Meters- Smart Meter Model

From the figure 1 shows that a problem might exist when determining which numeric characters need to be recognized as the value measured by kilowatt meter as the representation of electricity usage. Preprocessing on the image is necessary due to the effects of non-homogenous illumination. The filtering technique is applied to enhance the image on the dataset. The segmentation method is proposed to localize the region of Interest (ROI). The recognition methodology consists of two main stages: Preprocessing and recognition as shown in Fig.2, various image preprocessing techniques are used to obtain the best feature extraction result when performing a principal component analysis for segmentation. The input image is loaded from a file and tested using an offline system designed as a prototype of the nonreal-time system.

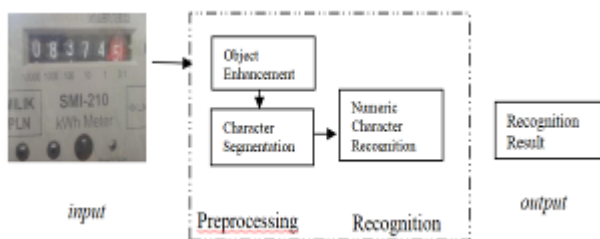


Figure 2. Diagram Block Recognition System

3.1. Preprocessing

The filtering technique is performed in this study for image enhancement [15]. Next step, the filtered RGB image converted to a binary image to reduce the variety of detected objects on image and focus on object morphology. This study performed a color space transformation method for KWh Meter image

enhancement before convert to binary number. RGB (Red-Green-Blue) Color Space that represents the color space of the power meter image is transformed into the HSV (Hue-Saturation-Value) color model. HSV color model is more suitable for image processing algorithm [16] where the colors are a natural color. Hue corresponds to pure color attribute, Saturation represents the degree of purity of color to white. Intensity is the size of the color brightness. After transformed into the HSV color model, histogram equalization is also performed for contrast enhancement. This method will sharpen the difference between background and the character on the image also among characters. Histogram equalization is a transformation from the distribution of intensity on the image, which can be performed by using formula Equation 4 below [14].

$$s_k = T(r_k) = \sum_{j=0}^k \frac{n_j}{n} = \sum_{j=0}^k P_r(r_j) \quad (4)$$

where r_k is the intensity level of the image. For the 8 bit grayscale image, it has 256 levels of color, then $r_k = 0$ to 255. Distribution of intensity r_k can be formulated as probabilities $P(r_k)$ of intensity on the n -pixels image. $P(r_k)$ is a portion n_k pixel number of r_k -intensity in the n -pixels image. The color image converted to grayscale color space before adjusted. Histogram equalization adjusts the original intensity value that lies between the initial value interval to the newly specified value range so as to increase the contrast. Fig. 3 illustrates the difference between KWh Meter image before and after enhancement.

The next step is Binarization that removes background image considered as noise for character segmentation as shown in Fig.3 (c). After binarization, numeral fields of the value measured by kilo wattmeter must be in a horizontal position and bounded in a rectangle box for segmentation as shown in Fig.5 (d).

The segmentation goal is to partition an input image into a region so that each one is as homogeneous as possible and neighboring ones are as different as possible. The character is localized by finding a region of interest (ROI) consist of pixels connected to one another as much as it can. PCA is performed on images which are the candidates of numeric characters. This step delivers segmented characters candidates of numeric character that will be recognized.

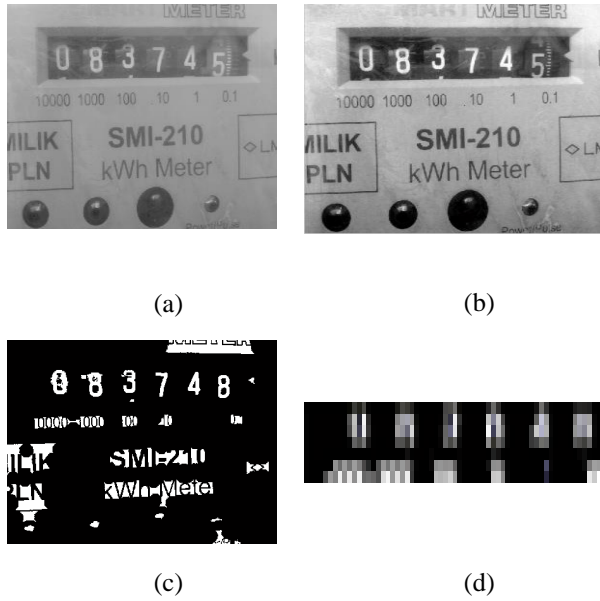


Figure 3. KWh Meter Image; (a) Before Enhancement, (b) After Enhancement, (c) Binarized image, (d) Localized Segmented image

3.2. Recognition

In this stage, Segmented images which are the candidates of numeric characters will be recognized by performing the matching process. The matching process uses Euclidean distance to measure its similarity. On Smart Meter Model, there six digits of numeric characters need to be recognized as the value measured by kWh Meter. In terms of effectiveness, recognition results accuracy on each image will be scored based on its six-character recognition performance.

4. RESULT AND DISCUSSION

The proposed approach evaluated by develop a prototype of the nonreal-time system in math software and programming, then tested using a dataset consisting of 50 images of the Smart Meter kWh Meter model. It is in JPEG format and 640 x 480 pixels size. Each image contains a numeric character that shows value measured by kWh Meter. These numeric characters have to be recognized using the proposed approach by performing PCA to extract its features for segmentation. Table-I summarizes the testing result of kilowatt-hour reading represents the amount of electricity usage. It shows accuracy in reading kWh meter is 90% on average. Miss recognition often happens for the last character of kWh-meter value when the character is not complete such as on ImageID12 in Table-I. The high similarity between characters, such as in ImageID2, numerical character 5 and 2 features components in the PCA transformed space also cause incorrect recognition. Fig.4 illustrates

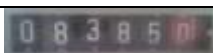

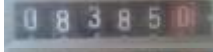
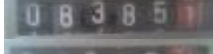

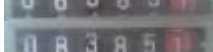

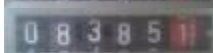

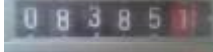
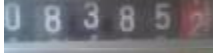
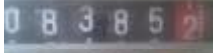
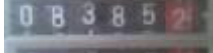
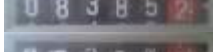
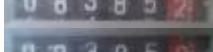

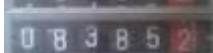

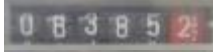
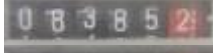
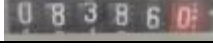


the character segmentation result of imageID2 that affects recognition performance.



Figure 4. Segmented characters

Table 1. Experimental Result

ImageID	Segmented ROI Image	Recognition	Accuracy
1		083374	100%
2		083742	83%
3		083745	100%
4		083745	100%
5		083748	100%
6		083748	100%
7		083751	100%
8		083751	100%
9		083751	100%
10		083837	83%
11		083835	100%
12		083832	83%
13		083840	100%
14		083840	100%
15		083840	100%
16		083847	100%
17		083846	100%
18		083841	83%
19		083847	100%
20		083847	100%
21		083847	100%
22		083848	100%
23		083848	100%
24		083848	100%
25		083848	100%
26		083849	100%
27		083850	100%

ImageID	Segmented ROI Image	Recognition	Accuracy
28		083850	100%
29		083850	100%
30		083850	100%
31		083851	100%
32		083851	100%
33		083851	100%
34		083851	100%
35		083851	100%
36		083851	100%
37		083851	100%
38		083851	100%
39		083852	100%
40		083852	100%
41		083852	100%
42		083852	100%
43		083852	100%
44		083852	100%
45		083852	100%
46		083852	100%
47		083852	100%
48		083852	100%
49		083852	100%
50		083860	100%

4.1. Conclusion

The task of automatic numeric character recognition is an important process of the smart grid in the electronic power industry. The automatic numeric character recognition problem on the electric meter image needs to be overcome. The results of the study proposed a recognition methodology for automatic numeric character recognition on the image by performing a principal component analysis for segmentation that has been proven its effectiveness in many cases. The experimental result shows its recognition accuracy score is 90% on average. It represents correctness in reading the value measured by kWh Meter. Although the algorithm which implemented works well for the task of action recognition, the

number of weaknesses are apparent. The image used in this study is taken from one angle shooting and does not cover a variety angle that might exist when the image is taken. Some improvements also needed in preprocessing for the future, especially the segmentation process that affects recognition results.

ACKNOWLEDGMENTS

We would like thanks to the Indonesian Ministry of Education and Culture and Indo Global Mandiri Foundation for the financial support of this research and future work.

REFERENCES

- [1] G. O. Young. 2009. Automatic Numeric Characters Recognition of Kilowatt Hour Meter. *Proceeding of the Fifth International Conference on Signal Image Technology and Internet Base System IEEE*.
- [2] M.R. Kim and D.S. Cho. 2013. The design of the data processing using AHP in automatic meter reading system. *International journal of computer science Issues*. 10, 1.
- [3] A. Sudiarso and R.J. Merischaputri. 2013. An automation of Electricity Usage Reading on Postpaid kWh Meter using Kohonen Type Artificial Neural Network. *International Journal of Mining, Metallurgy & Mechanical Engineering (IJMMME)*. 1, 4.
- [4] C.P. Edward V. 2013. Support Vector Machine Based Automatic Electric Meter Reading System. *IEEE International Conference on Computational Intelligence and Computing Research*.
- [5] H.Lei, P. Zhang, Q. Zeng, and X.Li. 2010. Numerical Recognition of Power Meter on A Handheld Terminal. *Proceeding of the Third International Symposium on Electronic Commerce and Security Workshop (ISECS)*.
- [6] D. Jabba, M. Rodriguez, G. Berdugo, M. Calle, M.A. Jimeno, and E.E. Zurek. 2012. A pattern recognition procedure for the identification of digital numbers in electrical meters using neural networks. *IEEE Symposium on Industrial Electronics and Applications (ISIEA)*.
- [7] M. Rodriguez, G. Berdugo, D. Jabba, M. Calle, and M. Jimeno, 2013. HD MR: a new algorithm for number recognition in electrical meters. *Turkish Journal of Electrical Engineering & Computer Sciences*. Turkey. 22 : 87-96.
- [8] S.Ali and M. Shah. 2010. Human Action Recognition in Video Using Kinematic Features and Multiple Instance Learning. *IEEE Trans. Pattern Analysis and Machine Intelligence*. 32, 2.
- [9] J.R. del solar, and P. Navarrete. 2005. Eigenspace-Based Face Recognition: A Comparative Study of Different Approaches. *IEEE Transaction on System Man and Cybernetic*. 16, 7 : 817-830.
- [10] B.S. Santosh, D.S.Reddy, M.S. Vardhan, S.Subhani. 2019. Heart Disease Prediction with PCA and SRP. *International Journal of Engineering and Advanced Technology*. 8, 4.
- [11] A.Chokka and K.S. Rani. 2019. PCA Based Regression Decision Tree Classification for

- Somatic Mutations. *International Journal of Engineering and Advanced Technology*. 8, 6.
- [12] Pentland, B. Moghaddam, T. Stanner. 2005. View Based and Modular Eigenspaces for Face Recognition. *Proceeding of the IEEE Computer Society International Conference on Computer Vision and Pattern Recognition*.
- [13] J.R. del Solar, and P. Navarette. 2005. Eigenspace based Face Recognition: A Comparative Study of Different Approaches. *IEEE Transaction on System, Man, and Cybernetic*. 16, 7.
- [14] J. Shlens. 2003. "A Tutorial on Principal Component Analysis, Derivation, Discussion, and Singular Value Decomposition". Princeton.
- [15] S.L.Eddins, R.E.Woods, and R.C. Gonzales. 2014. "Digital Image Processing Using Matlab". Prentice Hall.
- [16] Gonzales,R.C and Woods dan R.E Eddins. 2004. "Digital Image Processing". Upper Saddle River, NJ, Preatice Hall.