Object recognition using cognitive artificial intelligence’s knowledge growing system: A preliminary study

A D W Sumari\textsuperscript{1,2,*}, I N Syamsiana\textsuperscript{1}, D R H Putra\textsuperscript{1,3} and R A Asmara\textsuperscript{1,3}

\textsuperscript{1}Cognitive Artificial Intelligence Research Group (CAIRG), Department of Electrical Engineering, State Polytechnic of Malang, Jl. Soekarno-Hatta No. 9, Malang 65141, Indonesia
\textsuperscript{2}Faculty of Defense Technology, Indonesia Defense University, IPSC, Sentul, Bogor, Indonesia
\textsuperscript{3}Department of Information Technology, State Polytechnic of Malang, Jl. Soekarno-Hatta No. 9, Malang 65141, Indonesia

\*arwin.sumari@polinema.ac.id

Abstract. Object recognition has been a challenge for an intelligent system. There have been various approaches to develop such a system by utilizing machine learning especially which are based on neuron that is, neural network and deep learning. Common problems when using those approaches are the first one is dataset availability and the second one is the number of data. Lack of data causes neural-based approaches cannot be well operated, while a small number of data causes low accuracy results on the system. From another point of view, a considered-new technology from Cognitive Artificial Intelligence (CAI) perspective called as Knowledge Growing System (KGS) which may cope with such problems. With the capability to build its own knowledge from nothing, KGS is able to carry out recognition while developing its knowledge regarding the phenomenon it is trying to recognize. In this research, we showed KGS capability to perform object recognition as it is developing knowledge when interacting with such object directly. We did a benchmark on face recognition use-case with some common machine learning methods to show their performance on a small number of data, and KGS showed good results. With 100 feature-set from 5 persons’ face images, KGS achieves Degree of Certainty (DoC) as much as 80\% which is the system’s prediction accuracy that enables it recognizing the person based on that-moment data. Even though it is still lower compared to machine learning methods, but KGS shows advantages that it does not require high computational cost because it requires no training and no model development. In this research, we also showed that KGS enables the fast-deployment light-operated object recognition system.

1. Introduction
Object detection and recognition is a very easy task for humans since they were born. Humans can do this task every second, every day in their life. They can detect an object continued with recognizing it in just a very short time. Object detection and recognition can be carried out by using whether one of their sensory organs or the combination of two or more sensory organs. One of the easiest tasks that is the hardest one to an intelligent system is visual object detection and recognition which utilizes eyes. With eyes, humans can see, watch, and observe images and videos. They can easily extract information from such data and use the extracted information to build knowledge. They then can use the obtained
knowledge for making a decision and action. Detection, Recognition, and Identification (DRI) are parts of computer or machine vision, techniques that are utilized to enable computer or machine having humans’ vision capabilities, such as detecting, recognizing, and identifying using their vision sensory organ.

Identification is the highest level of human’s vision capability in discriminating or differentiating one object to others. From his original paper [1], Johnson divided it into four criteria, namely, Detection, Orientation, Recognition, and Identification (DORI). Other researcher uses Detection, Classification, Recognition, and Identification (DCRI) [2]. Humans can do all those activities easily, but the machine has to work very hard to do the same ones. Humans’ vision sensory organ identifies an object through features such as size, shape, weight, smell, texture, and color. In some cases, by using only one feature, human is able to obtain an inference based on his identification. In some other case, the human needs more features in order to get better inference so as to minimize error in making a decision. Detection, recognition, and identification are actually the steps that have to be carried out to get the most accurate inference. The machine is only an entity that works if a human gives it instructions and works on behalf of the human to carry out works. To make the machine performs human-like vision capability, it has to be fed with appropriate inputs that have to be well processed using a correctly-selected image or video processing. Therefore, the primary things that have to be prepared are entities that represent image or video called features as mentioned previously.

In the real world, sometimes we find new objects which we have no knowledge about them before. If we put ourselves as a machine learning-based system, we will try to find the most possible match with the objects that we have ever learnt before. But if we put ourselves as a system that has no knowledge about any object, how can we know what objects are those? What will we be doing to recognize them and how do we do it? For a machine learning-based system to obtain very good classification accuracy, it needs a considerable amount of storage and computation power [3] because huge past data is required for training before used. When such a system is tasked to recognize an object in that-moment without training in the first place, it will not be possible that it can work without knowledge that it should develop during training. To answer this question, there has been a new perspective in AI known as Knowledge Growing System (KGS) [4] that has capability to develop its own knowledge while interacting with the objects. In this paper, we share our preliminary study on the use of KGS for object recognition.

In order to deliver an easy-to-understand presentation, this paper is arranged as follow. A brief introduction regarding the aim of this paper has been already given in Section I. In Section II, we deliver a brief on image processing and feature extraction methods which can deliver appropriated inputs to KGS, and also a brief on KGS. An example of the use of KGS for object recognition with a use-case in face recognition will be delivered in Section III. The paper is concluded in Section IV with some concluding remarks and also ways forward to enhance KGS for object recognition.

2. A brief on related theories

2.1. Recognition as the result of prediction

Recognition in a very simple definition is the ability to acknowledge the type of a detected object. Humans cannot recognize an object if it does not exist. The existence of an object in an image or video can only be known if it is detected by eyes. Humans can recognize what the object is if he already has knowledge about it. But if not, he will try to approximate what the object is. What he is doing is to predict what type of object is. Therefore, it can be said that recognition is the result of the prediction. The quality of recognition depends on the accuracy of the prediction. Prediction aims at forecasting unobserved outcomes or future behavior [5] and can be performed in two types, namely, quantitatively and qualitatively [6]. In the AI world, prediction can be using various methods whether neuron-based ones such as Neural Networks (NN) and deep learning, or statistical-based ones such as Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). On the other hand, there is also an alternative approach based on cognitive psychology called Knowledge Growing System (KGS) [7]. This is the primary reason that KGS is called as Cognitive AI [8].
2.2. **KGS as the recognition engine**

In AI learning, knowledge of objects can be obtained from two means. The first one is by giving labels to objects in supervised learning, and the second one is by creating clusters of objects which are differentiated by their attributes or characteristics in unsupervised learning. These kinds of learning require data at high volume in the first place to be used to train the AI methods. The less the volume of data, the less the accuracy of prediction. On the other side, KGS is designed to have the ability to use the data it acquires at that moment to detect and recognize objects. KGS adopts constructivism theory especially the theory of learning by interaction to enable brain generating knowledge. Interaction is performed by using its sensors. The quantitative value of the obtained knowledge will show the detected and recognized object. The detailed explanation of KGS can be found in Sumari and Ahmad [4] and Sumari et al. [9]. There are four main formulas for KGS as shown in (1), (2), (3), and (4).

\[
\lambda = \left(2^\delta - \delta\right) - 1
\]  \hspace{1cm} (1)

where \(\delta = 1, \ldots, i, \ldots, n\) is the number of sensors, or we can generalize it as the number of information sources, and \(\lambda = 1, \ldots, j, \ldots, m\) is a collection of knowledge generated as the result of the fusion of information from all relevant corresponding sensors.

\[
P(\psi^l_j) = \frac{\sum_{i=1}^{\delta} P(\theta^l_i)}{\delta}
\]  \hspace{1cm} (2)

\(P(\psi^l_j) \in \Psi\) and it is called as New Knowledge Probability Distribution (NKPD) which is the fusion result of the relevant corresponding sensors, \(P(\theta^l_i)\). This is a collection of information that can be furthered extracted to obtain inference or new knowledge. The inference or new knowledge at this point can be obtained by applying (3) which becomes the quantitative value of the grown knowledge. The accuracy of prediction can be measured with Degree of Certainty (DoC) by using (4).

\[
P(\psi^l_j)_{\text{estimate}} = \bigcirc \left[ P(\psi^l_j) \right]
\]  \hspace{1cm} (3)

\[
\text{DoC} = \left| P(\theta^l_{\text{estimate}} - \phi^l_j \right| \times 100\%
\]  \hspace{1cm} (4)

2.3. **Image processing and feature extraction**

Image processing is a method for processing images into digital forms for specific purposes. The inputs to image processing are images and the outputs are images with improved quality. For example, images that are less sharp in color, noise (such as white spots), and blurry (blurred) require a process to improve their quality. Therefore, images that have been performed with image processing will have better information [10,11]. Feature extraction is a way to get a value from the image, and aims to look for areas of significant features in an image depending on the intrinsic characteristics and applications. The region can be defined in a global or local environment and is distinguished by shape, color, texture, size, intensity, geometry, statistical properties, and so on [12-14].

One of the objects which has long been researched is the face, that is, how to recognize a person from his/her attributes. Before an object can be recognized, it has to go to image detection in the first place. There has been various model for image detection for human face. There is one model that has been a leading player in face detection arena, namely Viola-Jones which provides a fast and accurate object detection system like face region and facial features like eye, nose, and mouth [14]. After the face area and the face parts are successfully detected by the Viola-Jones model, the midpoint of the area is taken by using (5) [15]. This process is followed by the calculation of the distance of the face parts by using Euclidean Distance (ED) in (6). The results of all processes are the features of faces that become the inputs to KGS. The feature consists of the distance from the right eye to the left eye, the distance from the right eye to the nose, the distance from the left eye to the nose, the distance from the right eye with the mouth, left eye with nose and nose with the mouth.
\[(x, y) = \left( x_1 + \frac{\text{width}}{2}, y_1 + \frac{\text{height}}{2} \right) \]  

(5)

\[ED = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \]  

(6)

3. Object recognition using knowledge growing system

Object recognition using an already-trained AI system has been a common thing. Recognizing an object is done by carrying out the prediction. The prediction is obtained after comparing it with the knowledge about objects the system had been trained. It is not a hard task because the hard challenge is already handled during data collection for system training. For KGS, the challenge occurs during the interaction with the object. The more the attributes it can get, the more accurate the result will be. To observe KGS performance to recognize objects from that moment data, data of faces from five different persons will be used. The object recognition using KGS is illustrated in Figure 1.

![Figure 1. The illustration of object recognition with KGS.](image)

3.1. Data preparation and processing

We prepared objects, that is, face images from 5 different persons numbered 0 to 4. By using Viola-Jones, we obtained extracted features of 5 different person’s face which consist of facial features namely, distances between eyes and nose, distances between eyes and mouth, and distance between nose and mouth. Each individual person has 20 facial images taken from the same distance of 1.5 meters with the conditions as follows: the same lighting, the same taken time, almost the same pose, and almost the same expression. In this case, images with index numbered 1 to 20 belong to person 0, 21 to 40 belong to person 1, and so on. The extracted data is stored into a table using a spreadsheet application and then is converted to CSV form for further processing using KGS. A partial result of the five individual person faces feature-set is shown in Figure 3. As be seen in the two tables, there are no label for each person. An example of feature extraction for person 4 at index 2 is depicted in Figure 2.

![Figure 2. An example of face feature extraction by using Viola-Jones model for person 4.](image)
3.2. Recognition

We have developed KGS software for another purpose but because it is multi-purpose, so we used it for this experiment. According to the mechanism illustrated in Figure 2, we have obtained the results from KGS as depicted in Figure 4. The results show that indices 1 to 40 have DoC = 100%, indices 41 to 60 have DoC around 31%, indices 61 to 80 have DoC = 0%, and indices 81 to 100 have DoC around 88%. From these results, KGS recognizes that there are two same persons, that is, person 0 dan person 1 (indices 1 to 20 and indices 21 to 40), and three other different persons.

3.3. Analysis

KGS is able to recognize 4 out of 5 persons by finding DoC for each group of features in feature-set. KGS recognizes person 0 dan person 1 as the same person, while person 2 and person 3 as other ones. But for person 4 even though KGS is able to recognize, there are many anomalies and some indices have the same values. If the images were not informed at the first place came from 5 different persons where each person has 20 images and was inputted to the system in a series, then KGS may recognize the person 4’s pattern of NKPD as many persons. The pattern shown by indices 1 to 40 is consistent even though there is one index that is not consistent. The same thing occurs to indices 41 to 60. The most consistent pattern is shown by indices 61 to 80, and it is warranted to this pattern represents different person than the previous ones.

A benchmark with machine learning methods such as NN, K-NN, Bayes, and SVM. For machine learning methods, the benchmark was carried out by splitting the feature-set into training set and testing set with 80:20 ratio, where 80 features were used as the training set and 20 features were used as the testing set. For KGS, because it does not need training phase so the feature-set which contains 100 features was inputted to the system. The benchmark results are given in Figure 5. The ability of KGS to recognize 4 out of 5 persons without training can be considered as a good result. The problem why KGS recognizes person 0 as the same as person 1 or vice versa, is because the features of the two persons’
face images are very similar. This is a challenge to have much better feature extractor to be able to obtain different features for very similar images.

![Table](image)

**Figure 5.** Benchmark of the recognition results of KGS and machine learning.

### 4. Concluding remarks

#### 4.1. Conclusions

Object recognition is easy for human even he/she has not had knowledge about it in the first place. He/she easily recognizes it by having knowledge of its attributes. If this mechanism can be emulated in the machine, then we can have a human-like system which is fast-deployed and light-operated. In this study we utilized and used KGS as such system with person face images as the objects to be recognized. The experiment results as well as its benchmark results to machine learning methods show a good prospect for KGS as an object recognition engine in a recognition system. As shown in Figure 5, KGS without training achieves 80% accuracy results. This results however, is lower compared to machine learning methods. On the other hand, KGS does not require high computational cost because it does not need training and require to develop the model that has to be done for machine learning. The advantages of KGS over machine learning are it does not need past data and automatically cuts the training phase which impacts on fast-deployment. It also does need huge storage to store past data which impacts to light-operation in the context of computing infrastructure.

#### 4.2. Ways forward

KGS is a new technology and is a new perspective [16] as well as a novel field in AI [17]. However, there are still many things that can be explored to make it much better and mature for object recognition. There are things that will be carried out further as follows.

- Enhancing the recognition mechanism in order to cope with features with very small differences.
- Using different objects for next studies such as air, land, and sea vehicles.
- Developing the mechanism for object identification which is the highest level in DCRI level.

### References

[1] Sjaardema T A, Smith C S and Birch G C 2015 History and Evolution of the Johnson Criteria SANDIA Report SAND2015-6368 (Albuquerque, New Mexico: Sandia National Laboratories)

[2] Israel S A and Irvine J M 2012 Heartbeat biometrics: a sensing system perspective International Journal of Cognitive Biometrics 1(1) 39-65

[3] Davies E R 2018 Computer Vision Principles, Algorithms, Applications, Learning Fifth Edition (London: Elsevier) p 371

[4] Sumari A D W and Ahmad A S 2018 Cognitive Artificial Intelligence: Concept and Applications for Humankind Intelligent System 23 55
[5] Bzdok D, Altman N and Krzywinski M 2018 Statistics versus machine learning Nat Methods 15 233–234
[6] Dobson A J 2013 Introduction to statistical modelling (New York: Springer)
[7] Sumari A D W, Ahmad A S, Wuryandari A I and Sembiring J 2010 Constructing brain-inspired knowledge-growing system: a review and a design concept Second International Conference on Distributed Framework and Applications 2010 95-102
[8] Sumari A D W and Ahmad A S 2017 Knowledge-Growing System: the origin of the Cognitive Artificial Intelligence Proceedings of the 6th International Conference on Electrical Engineering and Informatics 2017 1-7
[9] Sumari A D W, Sereati C O, Syamsiana I N, Wibisono M N, Abdulharis R and Putra D H P 2020 Cognitive artificial intelligence decision support system based on knowledge growing system International Journal of Advanced Science and Technology (IJASt) 29 3734-3743
[10] Kang S, Choi B and Jo D 2016 Faces detection method based on skin color modeling J. Syst. Archit. 64 100–109
[11] Akakin H Ç, Salah A A, Akarun L and Sankur B 2006 2D/3D facial feature extraction Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning 6064 60641D
[12] Hidayati K, Utomo A B and Wibawanto H 2018 Pengenalan wajah menggunakan metode local binary pattern (LBP) dan principal component analysis (PCA) untuk citra berkualitas buruk e-Proceeding Eng. 5 7928–7938
[13] Pongakkasira K and Bindemann M 2015 The shape of the face template: Geometric distortions of faces and their detection in natural scenes Vision Res. 109 99–106
[14] C Rahmad, R A Asmara, D R H Putra, I Dharma, H Darmono and I Muhiqqin 2020 Comparison of Viola-Jones Haar Cascade Classifier and Histogram of Oriented Gradients (HOG) for face detection IOP Conference Series: Materials Science and Engineering 732
[15] Sabri N, Henry J, Ibrahim Z, Ghazali N, Mangshor N N A, Johari N F M and Ibrahim S 2018 A Comparison of Face Detection Classifier using Facial Geometry Distance Measure 2018 9th IEEE Control and System Graduate Research Colloquium (ICSGRC) 116-120
[16] Sumari A D W, Ahmad A S, Wuryandari A I and Sembiring J 2009 Knowledge Growing System: a new perspective on Artificial Intelligence, Proc. of the 5th Int. Conf. on Information & Communication Technology and System 2009 217-222
[17] Sumari A D W, Ahmad A S, Wuryandari A I, Sembiring J and Sahlan F 2010 An introduction to Knowledge-Growing System: a novel field in Artificial Intelligence Jurnal Ilmiah Teknologi Informasi (JUTI) 8 11-20