IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 19 January 2023, accepted 17 February 2023, date of publication 27 February 2023, date of current version 2 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3249573

RESEARCH ARTICLE

Object Recognition for Humanoid Robots Using Full Hand Tactile Sensor

SOMCHAI POHTONGKAM AND JAKKREE SRINONCHAT, (Member, IEEE)
Department of Electronics and Telecommunication Engineering, Rajamangala University of Technology Thanyaburi, Khlong Luang, Pathum Thani 12110, Thailand

Corresponding author: Jakkree Srinonchat (jakkree_s@rmutt.ac.th)

This research was funded by the National Research Council of Thailand [grant No. 1245725 (N41D640012)], and this research has received funding support from the National Science, Research, and Innovation Fund (NSRF) via the Program Management Unit for Human Resources & Institutional Development, Research, and Innovation [grant number B04G640045].

ABSTRACT With the development of human-like robotic technologies, robots have been advanced to sense and recognize objects, which is one branch of artificial intelligence to mimic human reactions. Although high-performance computer technologies are used, the ability of object recognition by touch is still low due to the lack of proper sensors. Therefore, this research studied the object learning and recognition system through robot touches by developing an artificial sensory system acting as an electronic skin with tactile sensors. The Tactile Sensor is developed in this research, consisting of 15 Tactile Sensor Arrays and the palm's touchpoints. Furthermore, recognition analysis was developed on Bag of Word (BoW) and Convolution Neural Network (CNN) algorithms. With the BoW technique, using Support Vector Machine (SVM) as a classifier with Moment Analysis Descriptor (MA) provided the highest accuracy, showing more than 80.15% accuracy from five grasping of an object. With the CNN approach, InceptionNetV3 provided the highest accuracy of 98.28% from only one capture of an object.

INDEX TERMS Bag of feature, bag of word, CNN, robot hand, tactile object recognition, tactile sensor, transfer learning.

I. INTRODUCTION

The human sensory system comprises five organs: eyes, ears, nose, tongue, and skin. The development of these five sensory systems for the machine can be divided into five types: Image Sensor, Microphone, Electronic Nose, Electronic Tongue, and Electronic Skin. When the humanoid robot was developed, both the robot's anatomy [1] and the robot's nervous system [2]. The development of robot anatomy has evolved to have the structure of a robot-like human [1], especially the part of the robot hand that is currently being developed to be very similar to a human hand [3]. It has five finger structure, can hold objects, and performs many tasks similar to humans [4]. When the sensory system is applied to the robot system, the robot's hands are also applied to the electronic skin [5]. The tactile function of the robot hand mimics the functioning of the human sensory system. The robotic system has a tactile sensor as a transducer to convert pressure into electrical signals. Information obtained by touch, known as ''Pressure profile,'' was sent to the computer for processing. It has a working principle like a human being with a receptor as a sensor that receives pressure and sends it as a nerve impulse to the brain for analysis and prediction.

The development of artificial haptic perception for robots is a challenge according to the amount of research in this field is still small compared to visual image work. However, the importance and perception of objects' touch are essential. The operation of the robot is essential to hold an object. Therefore, evaluating objects, including shape, surface, and hardness, requires physical and sensory systems. Humans can only distinguish objects using tactile perception. However, today's robots lack this skill. It is mainly due to the lack of touch sensors suitable for robotic hands and analytical methods suitable for each sensor.

Therefore, this research develops object recognition for humanoid robots using full-hand tactile sensors, which achieves to carry and continues from our previous work [6].

The associate editor coordinating the review of this manuscript and approving it for publication was Zijian Zhang.

The BoW and CNN are employed to improve the system's efficiency.

II. THEORY AND RELATED WORK

The related work in object recognition by handling is quite diverse due to developments in several aspects of this subject, beginning with studies of human haptic behavior [7] from physiological and psychological studies to understand human haptic perception and unfamiliar objects surveys for recognition [8]. Other studies include humanoid robot hand system design to enable object handling [3], [4], [9], [10], control of the robot hand system of a humanoid robot [11], [12], tactile sensor array design for humanoids robot hand [5], [6], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32]. Methods for recognizing objects based on tactile image recognition [6], [15], [18], [20], [21], [22], [23], [24], [28], [29], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43] and studies on the application of tactile image recognition for humanoid robot hands [6], [15], [18], [20], [21], [22], [23], [24], [27], [28], [29], [31], [32].

A. TACTILE SENSOR FOR HUMANOID ROBOT HAND

In 1995 Lederman et al. [7] study of the anatomy and mechanics of the human hand was reported, which was the basis of humanoid robot development, addressing the object handle behavior of hands in terms of six parameters cylindrical grasp, tip, hook or snap, palmar, spherical grasp, and lateral. Park et al. [9] introduced the sequence of handling objects to design a humanoid robot hand is investigated and described the learning mechanisms in object recognition. The information occurs while touching an object, such as holding the object, shape, and surface. This learning method is similar when applied to robotic hand learning. The shape and surface of the object currently use a tactile sensor developed for use with the robot's hand. The development of sensors for the humanoid robot hand has continued, as shown in Table 1. However, it is installed on some parts of the robot hand, such as Fingertip [16], [19], [25], [26], [27], [29], Finger [14], [15], [21], [31], Palm [6], [24], [32], Fingertip and Palm [17], [20], [28], etc. Installing some sensors in the robot hand is very different from the real human hand. However, some studies install sensors across the robot's hand [22], [23], but the robotic hand does not have all five fingers [23]. The development of a sensor for a robotic hand equipped with a full palm and five-finger sensor was presented by Liu et al. [22]. However, there is still a lack of practical analysis of the results.

B. OBJECT RECOGNITION BASED ON TACTILE IMAGE

One of the essential developments of the humanoid robot hand is tactile object recognition. It has been reported in various research studies [6], [7], [16], [19], [21], [22], [23], [24], [25], [29], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45]. The important element for processing is the installation data of the tactile sensor array, so-called tactile image recognition, as summarized in Table 2.

TABLE 1. Humanoids robot hand with tactile sensor.

Year	Sensor Resolution (Pixels)	Number of Sensors (ea)	Application
2007 [14]	8 × 8, 9 × 8	15	Finger
2009 [15]	4 × 7	8	Finger
2010 [16]	12	5	Fingertip
2010 [17]	12	9	Fingertip, Palm
2011 [18]	8 × 8	2	Gripper
2011 [19]	12	32	Fingertip, Palm, Arm
2012 [20]	4 × 7, 4 × 6	6	Fingertip, Palm
2012 [21]	13 × 6, 14 × 6	6	Finger
2012 [22]	4 × 4, 4 × 3, 4 × 12, 5 × 9, 8 × 9	18	Hand
2014 [23]	1	241	Hand
2015 [24]	16 × 16	1	Palm
2015 [25]	1	5	Fingertip
2015 [26]	1 × 3, 1 × 5	4	Finger tip
2016 [27]	19	5	Fingertip
2017 [28]	3 × 8, 7 × 4	4	Fingertip, Palm
2018 [29]	19	5	Fingertip
2018 [30]	2 × 2, 3 × 6	7	Finger, Palm
2019 [31]	4 × 6, 4 × 4	15	Finger
2020 [32]	16 × 16	1	Palm
2021 [6]	16 × 16	1	Palm

A research group developed an algorithm for tactile sensors, which can be divided into two groups [44]: 1) a group that uses a small tactile sensor to test an object. If the sensor is much smaller than the object, we call this tactile image a "Local shape tactile image" [34], [35], [39]. The recognition method is called "Local shape recognition," which uses the Bag of Word (BoW) method or Bag of Feature (BoF) method for analysis. 2) The other group uses a tactile sensor that is large or similar in size to the object. We call this tactile image a "Global shape tactile image." This recognition method is called "Global shape recognition." It uses Machine Learning methods (ML) [15], [18], [20], [21], [22], [24], [36], [37], [38] and Deep Learning or Convolution Neural Network (CNN) methods for analysis [6], [28], [31], [32], [40], [41], [43]. The equations for the Local shape tactile image and Global shape tactile image are shown in (1) and (2).

Local shape tactile image

$$\text{Tactile sensor size} \ll \text{Object size} \tag{1}$$

Global shape tactile image

$$\text{Tactile sensor size} \approx \text{Object size} \tag{2}$$

The robot hand is equipped with sensors of various sizes in various positions. We can use two methods of recognition analysis. Because each sensor will give a small image, we can use the Local shape tactile image recognition method by using the BoW, and it can be analyzed by the Global shape tactile image recognition method by bringing the Local

TABLE 2. Recognition method for tactile image.

Year	Image size (mm^2)	Number of Grasps	Method	Recognition rate (%)
2009 [34]	6 × 14(2)	10	Vectorize, BoW	84.6
2009 [15]	4 × 7(8)	1	PCA, KNN	94.0
2011 [18]	8 × 8(2)	1	Mean, SD, KNN	92.0
2011 [35]	6 × 6	50	PF, PCA, BoW	81.4
2012 [20]	4 × 7(3), 4 × 6(3)	1	Moment, SVM	90.0
2012 [21]	13 × 6(3) 14 × 6(3)	1	PCA, SOM	93.0
2012 [22]	25 × 23	1	Segmentation, ANN	91.0
2012 [36]	24 × 16	1	Vectorize, PCA, KNN	81.4
2013 [37]	32 × 32	1	Haar Wavelet, KNN	86.0
2014 [23]	241	1	Vectorize, DNN	91.1
2014 [39]	10 × 10	1	FD, GE, MKL-SVM	85.5
2015 [24]	16 × 16	1	Vectorize, ANN	96.0
2015 [38]	6 × 14	15	SIFT, BoW	89.9
2016 [42]	6 × 14	20	iCLAP	85.36
2017 [28]	3 × 8(3), 7 × 4(1)	1	CNN	71.3
2017 [43]	768	1	AlexNet-DCNN	98.3
2017 [40]	28 × 50	1	CNN + SVM	91.6
2019 [31]	4 × 4(11) 4 × 5 (4)	1	CNN	95.5
2019 [41]	28 × 50	1	ResNet-CNN	95.3
2021 [32]	16 × 16	1	ResNet50-CNN	98.97
2021 [6]	16 × 16	1	InceptionResNetV2-CNN	91.8

shape tactile image obtained from different parts of the palm together to form a large image [13], [14], [15], [17], [20], [21], [22], [23], [28], [30], [31].

Therefore, we have used two analyses, the BoW, and the CNN methods, to compare the results.

III. MATERIAL AND METHOD

The materials and methods used in this study were divided into two main parts, hardware, and algorithms. The hardware used in the experiments consisted of a humanoid robot hand equipped with a tactile sensor array and the object sets used for testing in this study. The algorithm set was developed to control, operate, and collect all humanoid robot hands and analyze the tactile image for object recognition.

A. OBJECT SHAPE DETECTION BY TACTILE SENSOR

The humanoid robot development began with studying and copying actual human activity, for example, in the research on the anatomy and mechanics of the human hand [8], to design a humanoid robot [10]. Therefore, robot hand in this research is designed based on the human hand, as shown in Figure 1. There are 14 points of Degree of Freedom in the palm, consisting of 2 points on the thumb and 3 points on each other finger. Moreover, there are six types of hand behavior in humans who hold objects [8]: cylindrical grasp, tip, hook or snap, palmar, spherical grasp, and lateral. The cylindrical grasp, which grasps around the objects, is mainly applied in this research and can provide much information about grasping objects.

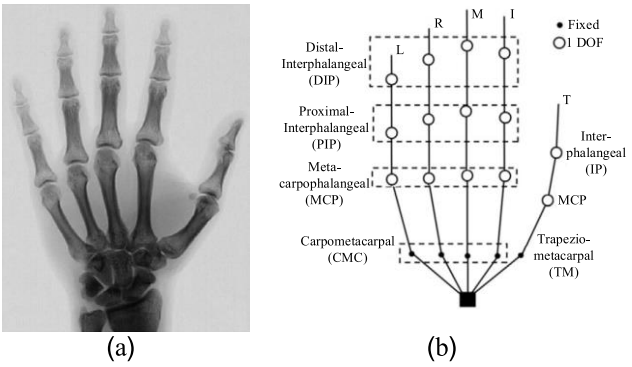


FIGURE 1. (a) Bones in a human hand, (b) hand's degree of freedom (DOF) [10].

Object shape detection using a tactile sensor relies on readings from the three-dimensional surface morphology of an object [46]. Also, the surface characteristics of an object can be obtained from the pressure that the object touches the tactile sensor, shown in Figure 2.

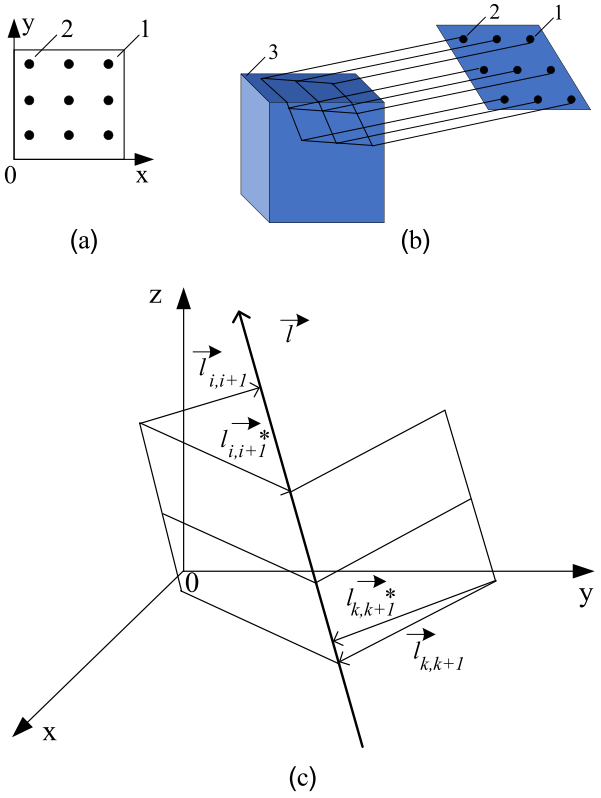


FIGURE 2. Touching effect with the tactile sensors (a) sensor array, (b) distance sensor array, (c) vector data of the fold surface object [46].

Define the sensor position on the x and y axis, as shown in Figure 2 (a). The force activates the sensor when the object is touched, as shown in Figure 2 (b). Moreover, Figure 2 (c) is the sensor data vector.

When the vector is

$$\overrightarrow{i_{i,i+1}} = [oy_{i+1} - y_i z_{i+1} - z_i]$$

and

$$\overrightarrow{i_{k,k+1}} = [oy_{k+1} - y_k z_{k+1} - z_k]$$

The slope k_i and k_k of vector $\overrightarrow{i_{i,i+1}}$ and $\overrightarrow{i_{k,k+1}}$ can be found as shown in (3) and (4).

$$k_i = \frac{z_{i+1} - z_i}{y_{i+1} - y_i} \tag{3}$$

$$k_k = \frac{z_{k+1} - z_i}{y_{k+1} - y_k} \tag{4}$$

The variance of s_i and $s_k$of set $\{k_i\}$and $\{k_k\}$can be found as shown in (5) and (6).

$$s_i = \sum_{i=1}^{n} \left(k_i - \overline{k_i}\right)^2 / n \tag{5}$$

$$s_k = \sum_{k=1}^{n} \left(k_k - \overline{k_k}\right)^2 / n \tag{6}$$

where $\overline{k_i}$ and $\overline{k_k}$ are the average set of $\{k_i\}$and $\{k_k\}$ as shown in (7) and (8).

$$\overline{k_i} = \sum_{i=1}^{n} k_i / n \tag{7}$$

$$\overline{k_k} = \sum_{k=1}^{n} k_k / n \tag{8}$$

The object characteristic can be analyzed from s_i, s_k, $\overline{k_i}$ and $\overline{k_k}$

1) In the case of the values of s_i, s_k are very small and $\rightarrow \overline{k_i} - \overline{k_k} 0$. It means the object's shape is the plan.
2) In the case of the value of s_i, s_k are very small and $\overline{k_i}$ - $\overline{k_k} \gg 0$). It means the object's shape is a fold surface of two planes.
3) In the case of the value of s_i, s_k are only one biggest and one smallest. It means the object's shape is a fold surface consisting of a plane and a curve surface.
4) In the case of the value of s_i, s_k are both significant values. It means the shape of the object is a curve surface.

However, the object is complicated, the touching areas have many patterns, and the objects' shape is directly affected by the touching objects. Therefore, the development of a humanoid robot hand is also based on the holding object position, as shown in Figure 3.

When the robot hand holds the objects, the sensor on the palm and fingers generates different signals, as shown in Figure 3. In Figure 3 (a), the robot hand holds the circular object, and the touching point occurs in every sensor. It differs from Figure 3 (b), in which the robot hand holds the rectangular objects. Therefore, the shape of objects can be analyzed with tactile images from the tactile sensor on the robot's hand. If the number of sensors increases, the tactile images are also increased. However, the shape analysis with the three-dimensional surface morphology [46] can not be used with a complicated shape, such as one object having various surface types.

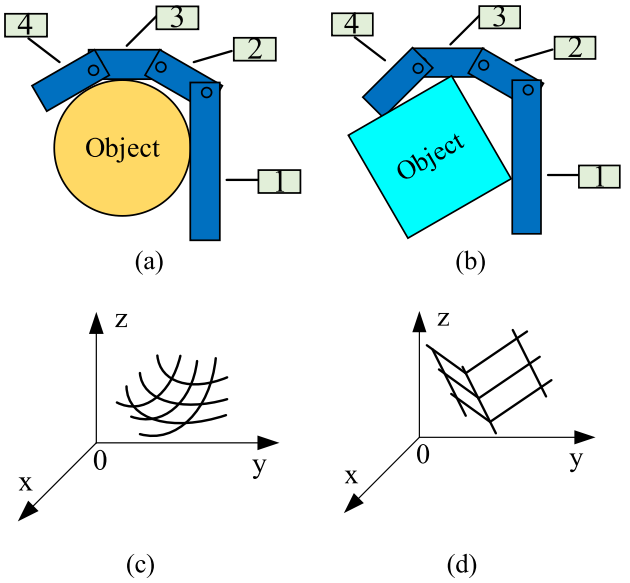


FIGURE 3. The holding pattern of the robot hand with the sensor on the palm and fingers (a) holding the circle objects (b) holding the rectangular objects (c) vector data of circle objects (d) vector data of the rectangular object.

B. FULL HAND TACTILE SENSOR AND ROBOT HAND

The tactile sensor used in this research was developed from a square electrode piezoresistive sensor [6], [45] by dividing the number of sensors into 15 parts according to the characteristics of the palm [8], [10]. The sensor is designed to be a Full Hand Tactile Sensor, as shown in Figure 1. The installation position is based on the structure of the fingers, divided into 3 types: 1) The palm tactile installation is a single position with a size 16 × 16 pixels or 56.0 × 56.0 mm^2 at the palm center (Metacarpals), 2) The position of the thumb by attaching 2 sensors along the knuckle, which installed at the position of distal phalanges by the size of 5 × 6 pixels or 17.5 × 21.0 mm^2 and installed at the position of proximal phalanges by the size of 5 × 5 pixels or 17.5 × 17.5 mm^2 and 3) finger position of 4 fingers with 3 sensors attached to the knuckle, position of Distal phalanges 4 × 6 pixels or 14.0 × 21.0 mm^2, Install at Intermediate phalanges 4 × 2 pixels or 14.0 × 7.0 mm^2 and Proximal phalanges 4 × 4 pixels or 14.0 × 14.0 mm^2. Therefore, there are 15 sensor installation points totaling 551 pixels. The entire sensor assembly is mounted onto a humanoid robot hand. The complete sensor assembly is shown in Figure 4.

The piezoresistive sensor used in this research was a conductive polymer (capLINQ, MVCF-40012BT50KS/2A).

TABLE 3. Tactile sensor for humanoid robot hand sensor.

Sensor	Number of Set	Resolution (Pixels)	Size (mm^2)	Total Pixel (Pixel)
Palm	1	16 × 16	56.0 × 56.0	256
Thump	1	5 × 6, 5 × 5	17.5 × 21.0 17.5 × 17.5	55
Finger	4	4 × 6 4 × 4 4 × 5	14.0 × 21.0 14.0 × 14.0 14.0 × 17.5	240

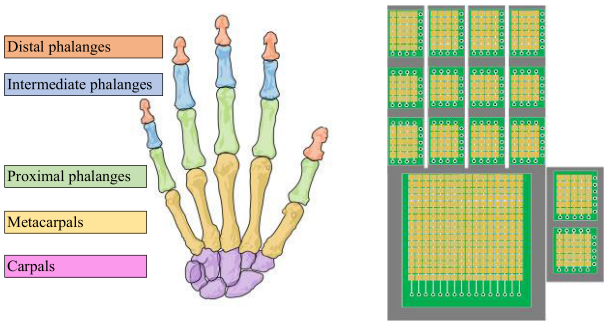


FIGURE 4. Design of sensors based on the location of the human hand, (a) the structure of the human hand [31] (b) the position of the designed sensor.

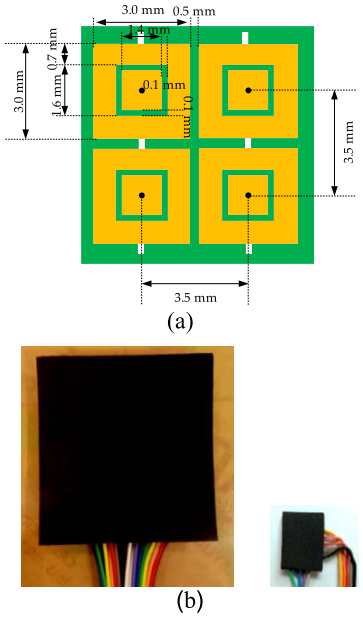


FIGURE 5. The sensor used in this research: (a) Sensor Electrode [6] (b) The actual sensor assembly.

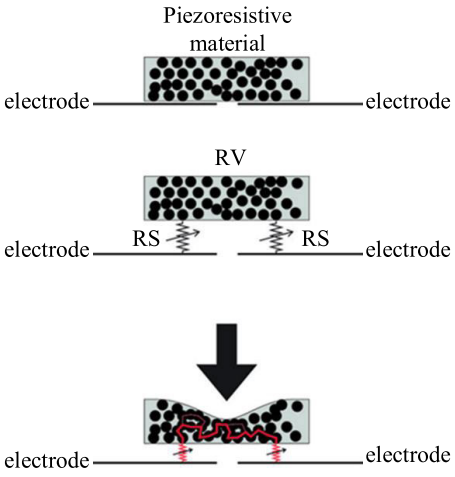


FIGURE 6. The working principle of piezoresistive tactile and resistive sensors [47].

The surface resistance of the conductive polymer is 50,000 ohm/cm^2. The PCB technology was applied to fabricate the sensor structure using epoxy PCB as a substrate with a thickness of 0.5 mm. The electrode made of Cu with a thickness of 0.2 mm was gold-plated at 18 μm. Sensors are divided into 6 sizes, as shown in Table 3. The electrode is designed using a pixel size of 3.5 × 3.5 mm^2. The external and internal size of the electrode was 3.0 × 3.0 mm^2 and 1.4 × 1.4 mm2, respectively. The gap between the internal and external electrodes was 0.1 mm, and the distance between each pixel was 0.5 mm, as shown in Figure 5 (a) [6]. An essential part of the Tactile Sensor Array is the elastic overlay layer. It is responsible for transmitting and distributing pressure to the Tactile Sensor Array. This research uses foam rubber with a thickness of 3 mm. A Tactile Sensor Array is obtained when all are assembled, as shown in Figure 5 (b).

In developing the tactile sensor, a square electrode design [6] was developed to reduce the dead area to a circuit electrode, and a piezoresistive tactile sensor was used because it has a simple structure. The reading circuit is simple and highly sensitive.

The developed piezoresistive tactile sensor was introduced by Karsten Weiß and Heinz Wörn in 2005 [47], as shown in Fig. 6. The piezoresistive layer is in contact with the rough surface electrode when a load is applied to the sensor. The piezoresistive will be pressed onto the surface of the electrode, and this compression directly increases the contact area, and it makes the reducing resistance between the sensor material and the electrode. The output resistance of the sensor consists of three resistors, as shown in Fig. 6. The RV is the resistance of the piezoresistive between the electrode, and the RS is the surface resistance that varies with the load applied. As shown in (9), the stress-strain relation is the relation between applied load F and the deformation ε of a linear elastic body, where E is the flexural modulus, and A is the body's cross-section area.

$$F = EA\varepsilon \tag{9}$$

Moreover, the R_S can be calculated as the pressure force, as shown in (10), in which the R_S is the surface resistance of the sensor material.

$$R_S(F) = \frac{1}{A(F)} \cdot R_{SO} \tag{10}$$

where the sensor resistance (R) is the total resistance of R_S and R_V as shown in (11)

$$R(F) = 2R_S + R_V \tag{11}$$

where R_V is the volume resistance of the sensor material over the electrode gap, which can be calculated from the piezoresistive characteristic, as shown in (12).

$$R_V = \frac{\rho \cdot K}{F} \tag{12}$$

where ρ is the resistivity of the piezoresistive, F is the force applied to the sensor, and K is a function of the elastic properties of the piezoresistive.

The tactile sensor is mainly developed in a square electrode type, which has the advantage of increasing the contact area over a circle electrode sensor because it can reduce the dead area [6]. Also, developing a piezoresistive sensor reading circuit is more accessible, which is the circuit's basis for measuring the current flowing through the sensor [19]. The developed sensor is calibrated the resistance value with a force gauge and an ohmmeter to measure the initial sensor resistance of the sensor. The sensor sensitivity is shown in Figure 7. The relation of the pressure and resistance can be defined as $R = aP^b$, where R and P are the pressure and resistance of the sensor, respectively. This relation can be calculated with (13).

$$R = 232 \times 10^3 P^{-0.97} \tag{13}$$

where R is resistance in Ω, and P is the acting pressure in kPa.

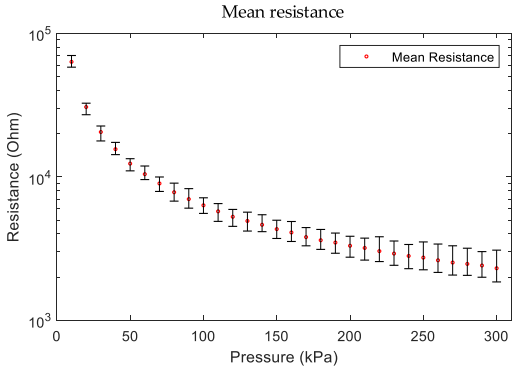


FIGURE 7. Sensor Sensitivity.

When the sensors are assembled according to Table 3, they are assembled into the robot hand. The 16 × 16 pixels sensors are in the palm, and a robotic hand suitable for this sensor is a flat palm. The robot's hand must have five fingers with 4 fingers except for the thumb, which must be divided into 3 parts according to the knuckle Sensor size 4 × 6 pixels, 4 × 4 pixels, and 4 × 5 pixels. The thumb has 2 parts, the size of 5 × 6 pixels and 5 × 5 pixels, respectively, as shown in Figure 8.

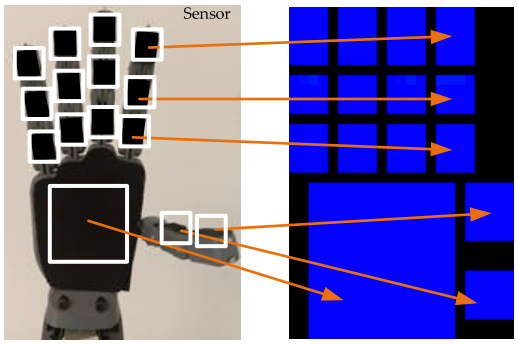


FIGURE 8. Installation of the sensor array to the robot hand and Tactile image pattern.

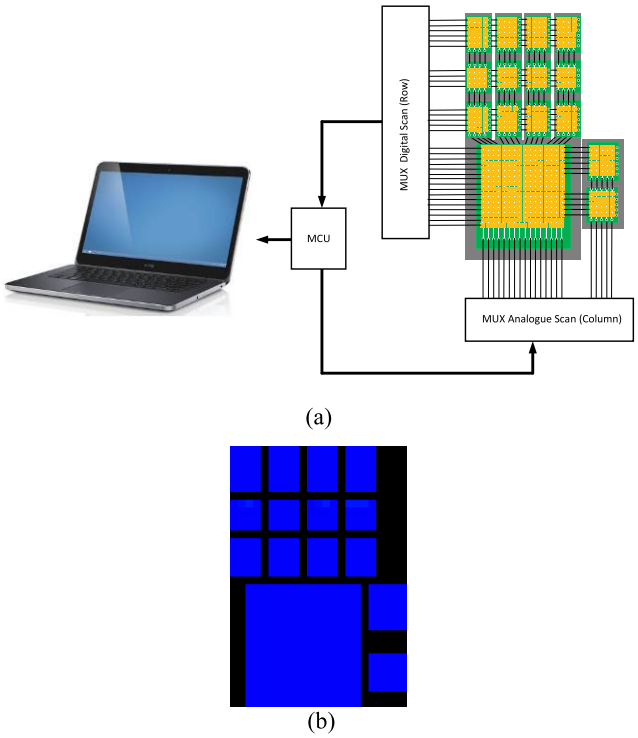


FIGURE 9. Tactile Image Acquisition Method (a) Toolkit diagram used for Tactile Image Acquisition (b) Tactile image pattern.

C. FULL HAND TACTILE SENSOR AND IMAGE ACQUISITION

Data acquisition uses Arduino microcontroller board mega 2560 model with Multiplexer number CD74HC4067. The work sequence is a Multiplex to scan the sensor in each position. Start by doing a multiplex in the Row position so that the output is logic "1" in Row that the system wants to read. ($R_i = 1$,) and logic "0" in other rows ($R_1, \ldots, R_{(i-1)}$, $R_{(i+1)}, \ldots, R_n = 0$). Then multiplex the column position to read the values in each sensor in the column starting from C_1 to C_m. The reading used A/D in the microcontroller, converted it to digital data, and transmitted it to the computer signal processing part. Scan the digital line in the next row

position when reading all columns until all sensors are read. The operation is shown in Figure 9.

The sensor touching area is much smaller than the object; therefore, the sensor sequence information consists of many surface areas of the object $O = \{o_1, o_2, \ldots, o_n\}$. When the full-hand tactile sensor, which contains 15 sensors, is used. The sensor provides information as $P = [p_1, p_2, \ldots, p_{15}]$ which is the sequence of images. It can be noticed in (14) and (15).

$$P = \begin{bmatrix} p_1 & 0 & p_2 & 0 & p_3 & 0 & p_4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_5 & 0 & p_6 & 0 & p_7 & 0 & p_8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ p_9 & 0 & p_{10} & 0 & p_{11} & 0 & p_{12} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & & 0 & p_{14} & & \\ & 0 & 0 & & 0 & 0 & & \\ & 0 & 0 & p_{13} & 0 & 0 & & \\ & 0 & 0 & & 0 & p_{15} & & \end{bmatrix} \tag{14}$$

$$p_x = \begin{bmatrix} s_{1,1} & s_{1,2} & s_{1,3} & \cdots & s_{1,n} \\ s_{2,1} & s_{2,2} & s_{2,3} & \cdots & s_{2,n} \\ s_{3,1} & s_{3,2} & s_{3,3} & \cdots & s_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{m,1} & s_{m,2} & s_{m,3} & \cdots & s_{m,n} \end{bmatrix} \tag{15}$$

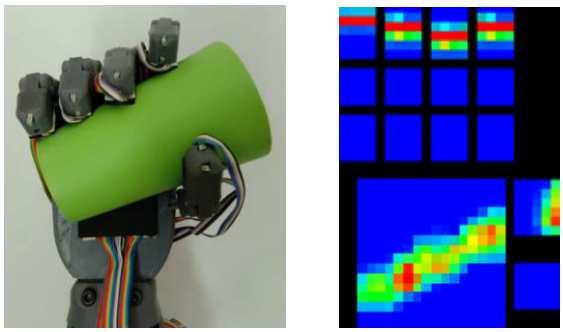
D. OBJECT EXPLORATION AND DATASET

The tactile images are obtained by touching different positions of the objects, while the grasping position is obtained by rotating an object on the 3D Coordinate. Once the robot has grasped an object, the physical characteristics of the robot's hand can freely manipulate the object, as shown in Figure 10. The experiment tests the proposed system with 20 objects or 20 classes. It consists of objects, as shown in Table 4. The dataset is created by grasping an object and saving 200 images per object, containing all 4000 images. The object is randomly positioned and rotated to obtain an image around the object. Therefore, it is an image obtained by capturing objects from different positions, as shown in Figure 11. It showed the tactile sample images of three classes of objects at different positions. (Five images per class).

The cylindrical grasp is applied chiefly to this proposed method to take the object information when grasping the object. The method of sample image collection from an object was carried out by random sampling of the object positions, and the images obtained from touching resulted from different object positions. The handle position resulted from rotating the object on 3D axes before entering the robot's hand and then allowing the physical characteristics of the robot's hand to handle the object freely to obtain tactile images representing the object at angles obtained randomly from the actual position. Figure 10 shows the sample object handle in this experiment, obtained from random object rotation in the XY-plane and along the Z-axis. The robot hand was allowed to handle the object, and tactile images were obtained from random positions and stored as a dataset for further testing.

TABLE 4. Dimensions of objects used in the model training.

Class	Object	Size
1	Battery	∅=33 mm, H=60 mm
2	Cup	∅=85 mm, W=120 mm, H=95 mm
3	Plastic tongs	L=165 mm, W=85 mm, H=15 mm
4	Remote controller	∅=33 mm, H=60 mm
5	Golf Ball	∅=40 mm
6	Scissors	L=18 mm, W=80 mm, H=12 mm
7	Variable wrench	L=200 mm, W=60 mm, H=14 mm
8	Allen key	L=90 mm, W=35 mm, H=6 mm
9	Screwdriver	∅=38 mm, L=220 mm
10	Tape measure	∅=70 mm, W=92 mm, H=47 mm
11	Cube	L=25 mm, W=25 mm, H=25 mm
12	Rectangular	L=50 mm, W=25 mm, H=25 mm
13	Triangular prism	L=50 mm, W=25 mm, H=20 mm
14	Pyramid	L=25 mm, W=25 mm, H=50 mm
15	Tetrahedron	L=25 mm, W=25 mm, H=50 mm
16	Large cylinder	∅=120 mm, L=110 mm
17	Small cylinder	∅=25 mm, L=50 mm
18	Cone	∅=25 mm, L=50 mm
19	Oval	L=25 mm, W=50 mm
20	Sphere	∅=25 mm

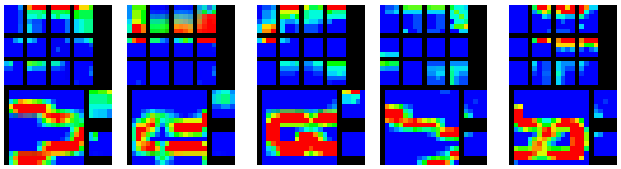


(a) (b)

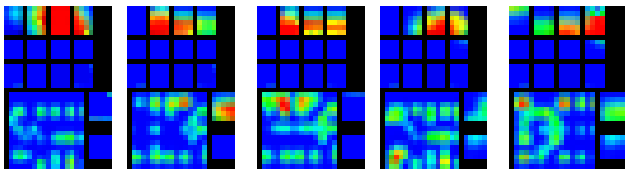
FIGURE 10. (a) handle the object of the robot hand (b) handle the object of the robot hand using the tactile image.

E. TACTILE OBJECT RECOGNITION METHOD

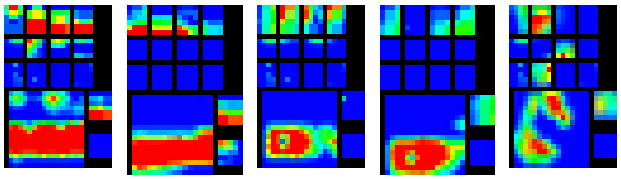
When grasping an object by hand, tactile image data is obtained from 2 main parts; the small image data is obtained from each finger part and the other image from the palm. In each image, there is a Local Shape Image; grabbing an object has the nature of grasping an object around the object. When arranging all the resulting image data, it will be a Global Shape Image. Therefore, the methods for analyzing the recognition can be done by the Local Shape Image Recognition method using the BoW and Global Shape Image Recognition models. CNN was used in this work, so both methods were used to compare performance.



(a) Sample tactile images from the "Plastic tongs" object.



(b) Sample tactile images from the "remote controller" object.



(c) Sample tactile images from the "Variable wrench" object.

FIGURE 11. Sample tactile images from a dataset created for the experiment. (a) Tactile image from the "Plastic tongs" class, (b) Tactile image from the "remote controller" class, and (c) Tactile image from the "Variable wrench" class.

F. BAG OF WORD METHOD

The first method uses a BoW Model [34] with a workflow, as shown in Figure 12 [40], [45]. The workflow starts by dividing the dataset into 2 sets: Training Set and Testing Set. The Training Set teaches the model by clustering object properties using K-mean Clustering and creating a histogram representing each object class. Define the Cluster for K-Mean to 30 clusters. Then build the Histogram to train to a classifier. In this research, we tested with 3 classifiers: K Nearest Neighbors (KNN), Support Vector Machine (SVM), and Artificial Neural Network (ANN). In the model of KNN, K = 10.

Clustering for BoW was applied to image recognition and tactile object recognition [34], [35], [39]. Use the K-Mean method as an unsupervised learning method for grouping data. K-means perform clustering, which is based on a statistical basis. The number of data groups K was set and was grouped in the sample sets $\{x_1, x_2, x_3, \ldots, x_n\}$ when $x_i \in \Re^d$ The grouping method organized similar data into the same group. The clustering result was a group of samples and the centroid, which usually represented the group and was centered and created functions. The objective of this group organization was to identify the minimum value of the sample distance and the centroid of each group c_k was as follows in (16):

$$J(r, c) = \sum_{i=1}^{n} \sum_{k=1}^{K} r_{ik} \|x_i - c_k\|^2 \tag{16}$$

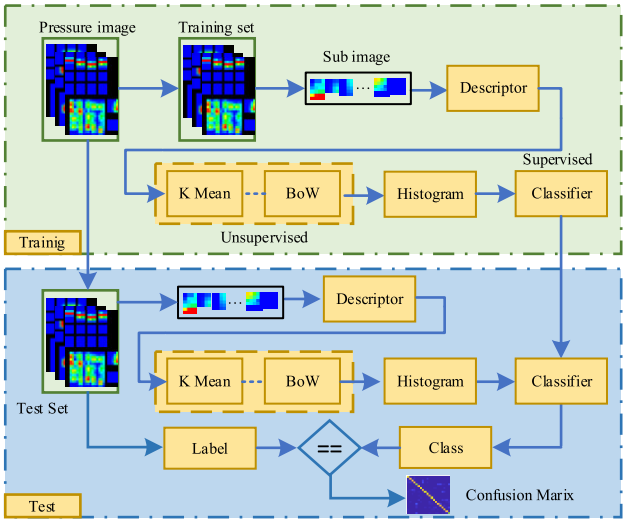


FIGURE 12. Work Flow of used BoW for object recognition [40], [45].

when $r_{ik} \in \{0, 1\}$ is a variable that tells the members of the group k of the sample i. when x_i was assigned to the group k has a value of 1 and will have a value of 0 for another group.

Three classifier methods were tested in the experiment: KNN, SVM, and ANN. For each classifier, three descriptors were tested: Moment Analysis (MA) [20], Scale Invariant Feature Transform (SIFT) [38], and Pola Furrier (PF) [35]. The experiment was tested on 30 clusters. The results were obtained from the test reported by the Classifier method. All tests were performed in 10 iterations, and average accuracy was determined. The operation is shown as Work Flow in Figure 12. The BoW method was used to test object recognition from a local shape tactile image, which is an image obtained by mounting a small tactile sensor on a finger. Therefore, it has presented a method for segmenting the image into sub-images divided into 14 finger images. The size of the image is according to the size of the sensor, which is 4 × 6 pixels image size 4 images, 4 × 4 pixels number 4 images, 4 × 5 pixels size 4 images, 5 × 6 pixels size 1 image, size 5 ×5 pixels amount of 1 image. The image from the palm position sensor has been divided into 16 sub-images and used the overlapping technique to maintain the suitable image size calculation. There is a partial overlap of the image, as shown in Figure 13. In the experiment, the image was divided into 4 × 4 sub-images, Overlapping into 16 sub-images (sub-images of 8 × 8 pixels). Therefore, there will be 30 sub-images used to make a descriptor for creating a data vector to represent the image obtained by capturing an object once.

G. BAG OF WORD TRAINING PROCESS

In order to get the object's information, it needs to touch many times on the object. The information of each touch can be defined as $P = [p_1, p_2, \ldots, p_{15}]$. The Bag of Words (BoW) is applied to analyze the object, as shown in Figure 14. The proposed process is as follows:

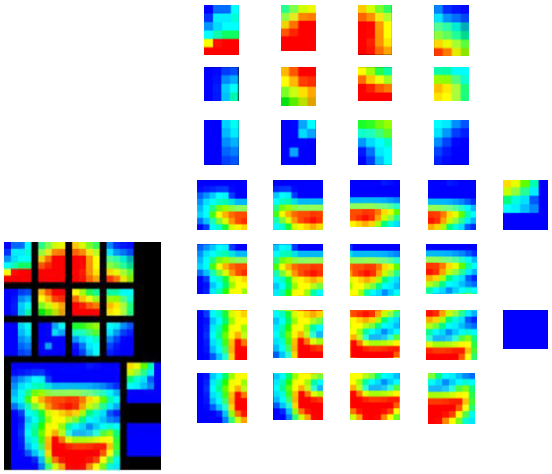


FIGURE 13. Dividing a Sub Image of a Full Hand Tactile Image for a Descriptor (a) Full Hand Tactile Image and (b) Sub image.

1) The tactile image data is collected using the full-hand tactile sensor. This image data is divided into the training and testing data set.

2) An image is divided into 30 sub-image as $P = [p_1, p_2, \ldots, p_{30}]$ and then the feature extraction is employed using MA, SIFT, and PF techniques. The results provide a descriptor as $W = [F_1, F_2, F_3, \ldots, F_{30}]$, and also the feature vector is given as $F_x = [f_{x,1}, f_{x,2}, f_{x,3}, \ldots, f_{x,n}]$.

3) Then, the descriptors operate into a vector quantization technique using k-means clustering with 30 clusters. The "word" is defined to represent each data of cluster.

4) The histograms of "word" in each cluster are created.

5) Finally, the histogram of training and testing is compared to find the winner using KNN, SVM, and ANN.

H. CNN METHOD

The second method used a CNN network with a diagram, as in Figure 15. The dataset was divided into 2 sets as the training data for network learning and then tested with testing data.

Convolutional neural networks are an architecture of feed-forward neural networks developed for image-learning purposes. In 1998, LeCun et al. [48] presented a convolutional neural network with LeNet-5 that consists of three important layers:

1. Convolutional Layer Convolutional to do feature extraction or extract important features of the image by calculating the dot product between the image sub-region and a filter or a kernel that looks like a square matrix, such as size 3 × 3. The result will be a new data set generated from h∗F is obtained. The newly created data set is called the Feature map. Convolution equation as shown in (17).

$$G[i,j] = \sum_{u=-k}^{k} \sum_{v=-k}^{k} h[u,v] F[i-u, j-v] \tag{17}$$

where h is the image matrix,

F is Kernel matrix of size $k * k$

2. Pooling Layer It is the layer next to the convolution layer. The function of this layer is to reduce the size of the feature map to a smaller size for sensitivity in calculations. The maximum calculation method is max pooling, where the feature map is divided into sub-region, the size is P × P, such as 2 × 2, then find the most significant value in each sub-region to create a new feature map.

3. Fully-Connected Layer is the network's last layer that is part of the Multilayer Perceptron (MLP) neural network classification, where the feature map in the last layer is the input imported into the neural network. The result obtained from the fully connected layer is an output prediction calculated by the softmax function to convert the prediction to the probability form (18).

$$\sigma(j) = \frac{exp(z_j)}{\sum_{k=1}^{K} exp(z_k)} \tag{18}$$

where σ is a softmax, z is an input vector, e^{zi} is a standard exponential function for the input vector, K is a number of classes in the multiclass classifier, and e^{zj} is a standard exponential function for the output vector.

In this work, we tested all 7 CNN models that were used in previous research [6], including AlexNet [49], VggNet19 [50], GoogleNet [51], InceptionNetV3 [52], InceptionRestNetV2 [53], XceptionNet [54], NASNetMobile [55], all of them are the most widely used CNNs in Computer Vision. They have been used to recognize objects from tactile images, providing high performance [6], [31], [32], [40], [41], [43]. A full-hand tactile sensor has been applied to object recognition using the Transfer Learning method [56]. To test the dataset with CNN, training parameters as in Table 5. The appropriate initial learning rate of each network is shown in Table 6. Each network model was tested 10 times, and the average accuracy was determined to compare the results.

TABLE 5. Training parameter.

Parameters	Values
NumEpochs	30
NumBatchSize	8
Momentum	0.9
LearnRateDropFactor	0.1
LearnRateDropPeriod	8

The object recognition using the full hand tactile image is operated with 4,000 images within 20 classes, and these 2,000 images are separated into training and testing data sets each. Then the training data set is passed to each CNN model, and then each CNN model is tested with the testing data set, as shown in Figure 16. The process can be explained as follow.

Pre-image processing – Each CNN model requests a different input image size according to using the Transfer learning in the testing step. Therefore, pre-image processing is

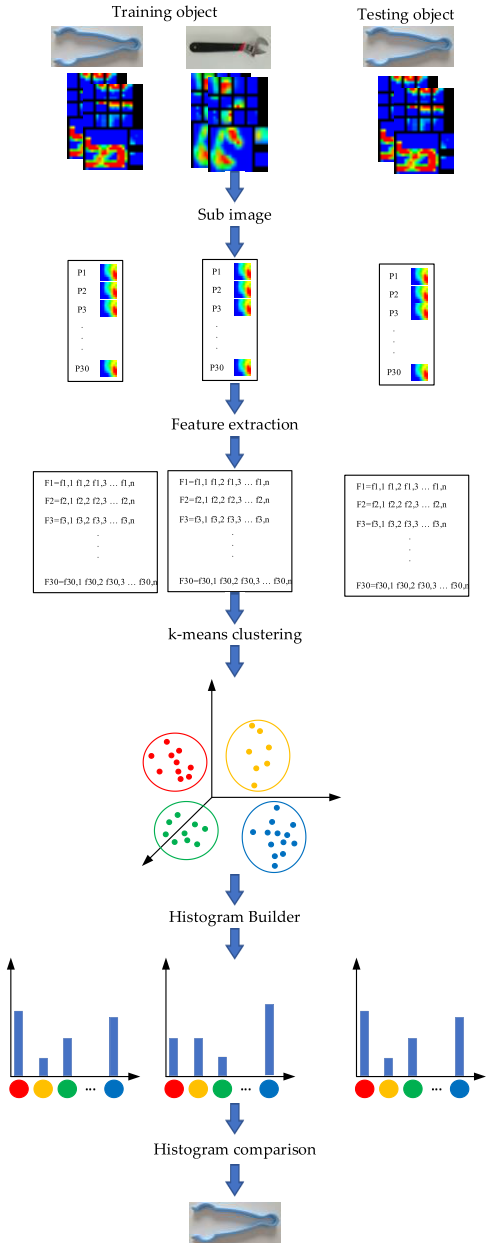


FIGURE 14. Recognition Analysis Process using BoW.

TABLE 6. Initial learning rate.

Network	Values
AlexNet	0.001
VggNet19	0.0001
GoogleNet	0.001
InceptionNetV3	0.01
InceptionRestNetV2	0.01
XceptionNet	0.01
NASNetMobile	0.01

appled to adjust the input image size to be suitable for each CNN model such as AlexNet, Vgg19, GooLeNet, and

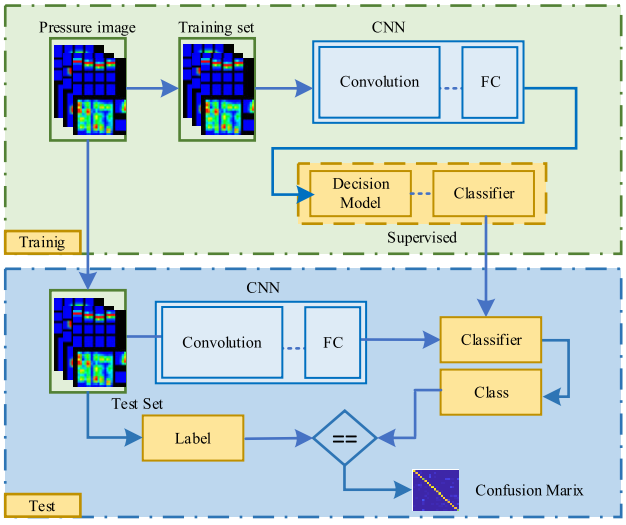


FIGURE 15. Work Flow of used CNN for object recognition [40], [45].

NASNetMobile uses 224 × 224 pixels, respectively InceptionNetV3, InceptionRestNetV2, XceptionNet uses 299 × 299 pixels.

Convolution layer – The feature extraction performs in this stage by the Dot Product calculation between the sub-region and filter or 3 × 3 kernel square matrix, creating a new data set called Feature Map.

Pooling Layer – This layer continues from the Convolution Layer, as shown in Figure 16. The pooling layer function is used for downsampling the feature map size using the max pooling or average pooling technique. The feature map is divided into the sub-region P × P pixels, such as 2 × 2, and the maximum values of each sub-region are calculated to create the new feature map to be the input of the next pooling layer. Each CNN model has a different number of pooling layers depending on the feature map size.

Dense Layer – This layer is a part of the multi-layer perceptron (MLP) structure classification. The feature map of the last pooling layer is activated by flattering and then pass it through the Artificial Neural Network (ANN). The output of the dense layer is then calculated by the softmax function, as shown in Equation (17), to predict the probability pattern.

IV. RESULT AND DISCUSSION

The experiment consisted of 2 parts: 1) the results of the recognition rate from the BoW model and 2) the results of the recognition rate from the CNN model, follows as:

A. RECOGNITION RESULT FROM BoW

In the BoW experiment, three classifiers were tested: KNN, SVM, and ANN. Each classifier was tested with Moment analysis, Pola furrier, and SIFT descriptors. There are 30 clusters. The results of the tests were divided by the classifier method, with every test performed 10 iterations, and then the mean was averaged. All test results are shown in Figure 17.

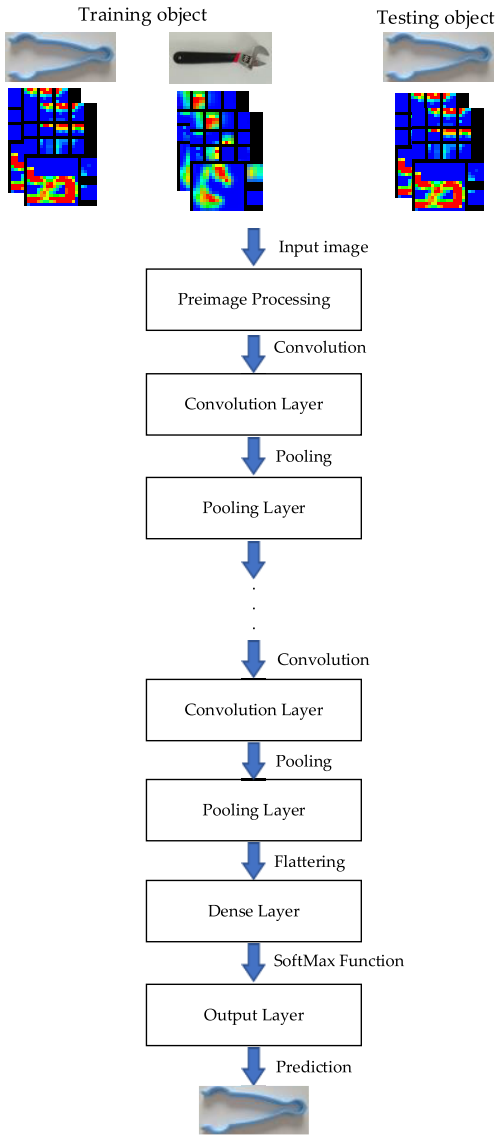


FIGURE 16. the process of analysis and recognition by the CNN technique.

The dataset obtained from the experiment was tested for recognition by the BoW method and shown in Figures 17 and 18. The test was performed with the classifier of 3 methods consisting of KNN, SVM, and ANN methods in Figure 17(a). The test results of the BoW method using the KNN method as a classifier by using all 3 methods of descriptor, compared to all descriptors, found that in the first period, the Moment analysis had higher accuracy than other methods. However, when holding the object up to 4 times onwards, it was found that the SIFT method gave higher accuracy.

The result of the SVM method is the classifier, as in Figure 17(b). the comparison results of all descriptors showed that the first period found the Moment analysis had higher accuracy than other methods, but when the object is kept up to four times, the SIFT method also provides higher accuracy. As for the result of the ANN classifier in Figure 17 (c)., it was found that the Moment analysis had higher accuracy than other methods in the first period. However, the SIFT method gives higher accuracy when holding the object up to four times. The test results showed that the BoW model characteristics have to grasp the object multiple times to predict the results accurately, and the classifier that gave the best test results was SVM, where the descriptor of the 3 methods had different recognition results. The Moment analysis descriptors are best at recognizing objects 1 to 4 as the first period, but when an object is grasped more times, the accuracy of each descriptor differs slightly. However, the SIFT method provides slightly higher accuracy than other methods.

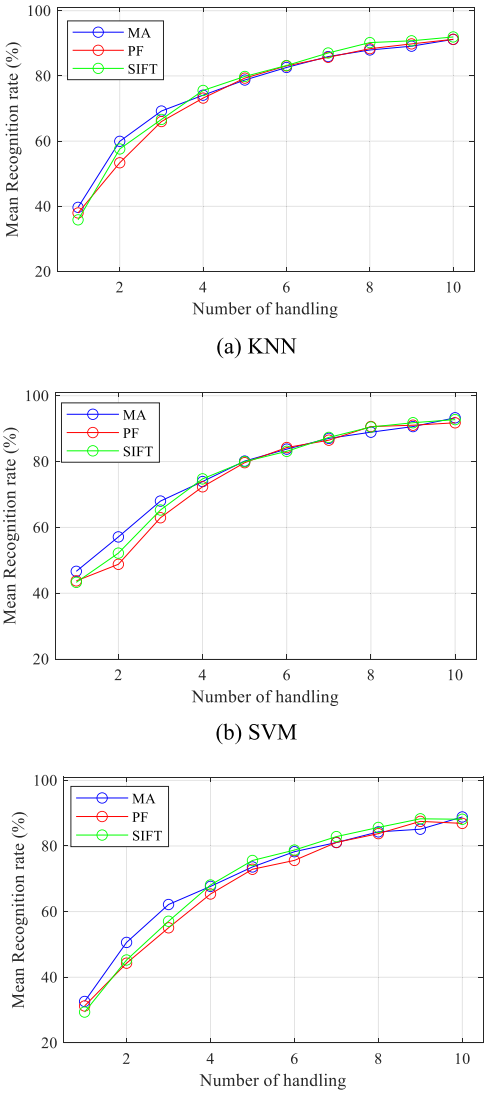
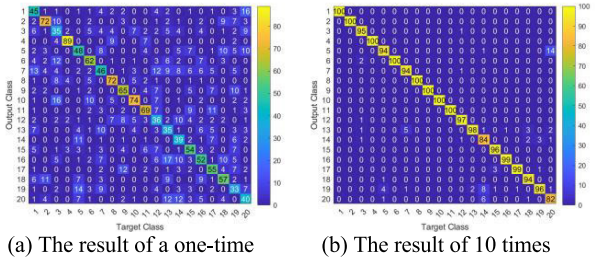


FIGURE 17. Recognition rate results were obtained from the Model BoW test using (a) KNN, (b) SVM, and (C) ANN.

B. RECOGNITION RESULT FROM CNN

The same data set as the BoW method was tested using the CNN 7 model consisting of AlexNet, VggNet19,



(a) The result of a one-time grasp (b) The result of 10 times grasp

FIGURE 18. The BoW object recognition results when the object is once grasped by the SVM method and SIFT descriptor (a) object is one-time grasp (b) object is 10 times grasp.

GoogleNet, InceptionNetV3, InceptionRestNetV2, XceptionNet, and NASNetMobile. The test results are shown in figure 19. The results showed that the highest prediction accuracy was InceptionNetV3, with an accuracy of 98.28%, and InceptionNetV3 gave similar results, with an accuracy of 98.26%. The results show that the CNN method is more accurate than the BoW method. Therefore, that is suitable for application to the recognition system of humanoid robots. In Figure 20., the recognition results of InceptionNetV3 show the recognition results in each class. It is found that the objects that the system has high efficiency are Cup, and a Tape measure, which has an accuracy is 100%. The class object with the lowest recognition was the Sphere because the tactile image of the Sphere was similar to a golf ball. It was predicted to be a golf ball, and it was found that many standard-shaped objects have more errors than everyday objects because many standard-shaped objects are similar. The comparison is in Figure 21. The error occurring with Class 5 (a golf ball) is related to Class 20 (Sphere) since the tactile image of a golf ball is similar to that of a Sphere. Moreover, it was found that AlexNet had significantly higher predictive errors in Class 5 and Class 20. Therefore, the accuracy is lower than the InceptionNetV3 method.

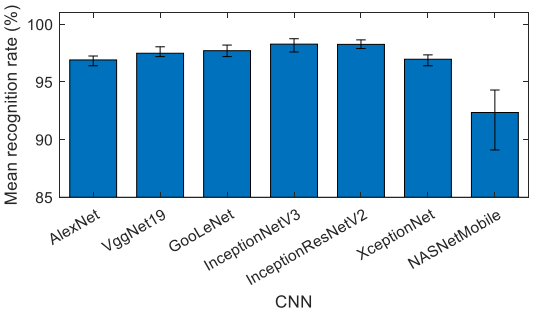


FIGURE 19. Comparison recognition rate from CNN.

Object Exploration to examine objects in detail by handling multiple objects grasps before prediction. It is a human-like behavior to be able to predict accurately. Therefore, the object survey was conducted several times by holding the object, using the CNN method to analyze it, and then determining the maximum probability (P_{max}) and summation of Probability (P_{sum}). The results are shown in Figure 22. by comparing the results with two analysis methods between the AlexNet and

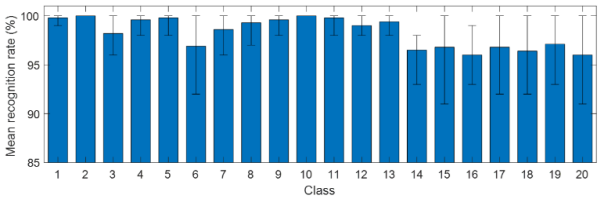
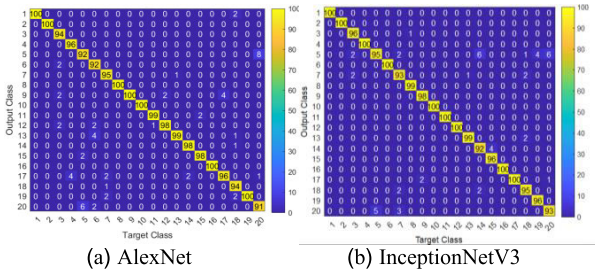


FIGURE 20. Recognition result for each class object from InceptionNetV3.



(a) AlexNet (b) InceptionNetV3

FIGURE 21. The confusion matrix of object recognition with CNN (a) AlexNet and (b) InceptionNetV3 methods.

InceptionNetV3. The results showed that increased object grasp would result in higher accuracy. When using the InceptionNetV3 method, the accuracy of 97.2% increased to 99.0% when holding objects increased to 2 times. The accuracy rate is stable and does not increase when the object is grasped more than 6 times, whereas, in InceptionNetV3, accuracy was 100% for more than 6 manipulations, and after about 10 times grasp prior to the prediction, both the maximum of probability and summation of probability methods were 100% accurate in the prediction.

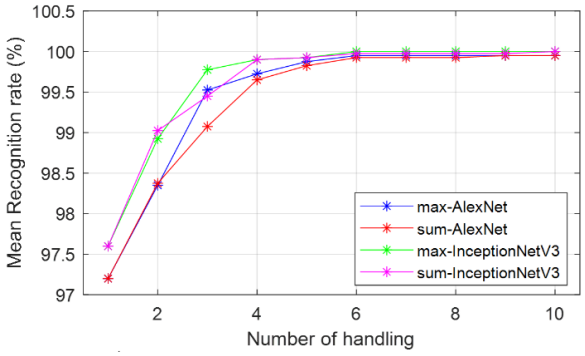


FIGURE 22. Results from Object Exploration.

In the efficiency test used in this study, DCNN training employed a CPU AMD Ryzen7 3750H 1200MHz and GPU (Nvidia GTX1650, US). The Bow and the CNN technique utilizes the CPU and GPU resources, respectively. The time consumption of each model is shown in Table 7. The results illustrated that the BoW technique used the time calculation at 0.113 – 0.208 s/image, and the BoW(KNN+SFIT) used the least time calculation at 0.113 s/image. On the other hand, the CNN indicated that it used the time calculation at 0.757 – 24.210 s/image, and the Alexnet used the least time calculation at 0.757 s/image due to its uncomplicated structure. The model that requires the most training time is VggNet19 (24.210 s/image), due to its more complicated

structure and more parameter and Network size. InceptionNetV3 requires 4.717 s/image, which is higher than the consumption time for AlexNet but provides a higher recognition rate.

TABLE 7. Training time.

Method	Training Time/image (S)
BoW (KNN+MA)	0.161
BoW (KNN+PF)	0.136
BoW (KNN+SFIT)	0.113
BoW (SVM+MA)	0.208
BoW (SVM+PF)	0.182
BoW (SVM+SIFT)	0.157
BoW (ANN+MA)	0.160
BoW (ANN+PF)	0.137
BoW (ANN+SIFT)	0.114
AlexNet	0.757
VggNet19	24.210
GoogleNet	1.837
InceptionNetV3	4.717
InceptionRestNetV2	13.927
XceptionNet	4.357
NASNetMobile	15.607

V. CONCLUSION

This research studies object recognition from image processing obtained by touch by developing a Full Hand Tactile Sensor. It consists of a tactile sensor array of 15 points along the touchpoints of the palm. Twenty objects were tested, and recognition was analyzed using BoW and CNN methods. The BoW method found that using SVM as a Classifier and Moment Analysis Descriptor provides the highest accuracy. It provides more than 80% accuracy in 5 object captures ($p = 5$) when clustering is divided into 30 clusters ($w = 30$). When using the CNN method was tested on the CNN 7 model. InceptionNetV3 Provides the highest accuracy, with 98.28% accuracy in a single capture. Therefore, all these developments are for object recognition based on touching objects obtained by processing the data read from the tactile sensor for further use in humanoid robots.

REFERENCES

[1] D. W. Seward, A. Bradshaw, and F. Margrave, "The anatomy of a humanoid robot," *Robotica*, vol. 14, no. 4, pp. 437–443, Jul. 1996, doi: 10.1017/S0263574700019846.

[2] T. Ogata and S. Sugano, "Mechanical systems for autonomic nervous system in robots," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatronics*, Tokyo, Japan, 1997, pp. 437–438, doi: 10.1109/AIM.1997.652983.

[3] L. D. Dunai, M. Novak, and I. L. Lengua, "Development of a prosthetic hand based on human hand anatomy," in *Proc. IECON 46th Annu. Conf. IEEE Ind. Electron. Soc.*, Oct. 2020, pp. 600–605, doi: 10.1109/IECON43393.2020.9254216.

[4] F. Negrello, H. S. Stuart, and M. G. Catalano, "Hands in the real world," *Frontiers Robot. AI*, vol. 6, pp. 1–13, Jan. 2020, doi: 10.3389/frobt.2019.00147.

[5] A. Saudabayev and H. A. Varol, "Sensors for robotic hands: A survey of state of the art," *IEEE Access*, vol. 3, pp. 1765–1782, 2015, doi: 10.1109/ACCESS.2015.2482543.

[6] S. Pohtongkam and J. Srinonchat, "Tactile object recognition for humanoid robots using new designed piezoresistive tactile sensor and DCNN," *Sensors*, vol. 21, no. 18, pp. 1–26, Sep. 2021, doi: 10.3390/s21186024.

[7] S. J. Lederman and R. L. Klatzky, "Extracting object properties through haptic exploration," *Acta Psychologica*, vol. 84, no. 1, pp. 29–40, Oct. 1993, doi: 10.1016/0001-6918(93)90070-8.

[8] C. L. Taylor and R. J. Schwarz, "The anatomy and mechanics of the human hand," *Artif. limbs*, vol. 2, no. 2, pp. 22–35, May 1955.

[9] J.-H. Park, J.-H. Bae, Y.-D. Shin, S.-W. Park, and M.-H. Baeg, "Framework of grasping planning for multi-fingered robot hands," in *Proc. 8th Int. Conf. Ubiquitous Robots Ambient Intell. (URAI)*, Incheon, Korea (South), Nov. 2011, pp. 455–458, doi: 10.1109/URAI.2011.6145863.

[10] L. Tian, N. M. Thalmann, D. Thalmann, and J. Zheng, "The making of a 3D-printed, cable-driven, single-model, lightweight humanoid robotic hand," *Frontiers Robot. AI*, vol. 4, pp. 1–12, Dec. 2017, doi: 10.3389/frobt.2017.00065.

[11] T. Tsuji, K. Shima, N. Bu, and O. Fukuda, "Biomimetic impedance control of an EMG-based robotic hand," in *Robot Manipulators Trends and Development*. London, U.K.: IntechOpen, 2010 https://www.intechopen.com/chapters/10475, doi: 10.5772/9184.

[12] D. Tanaka, T. Matsubara, and K. Sugimoto, "An optimal control approach for exploratory actions in active tactile object recognition," in *Proc. IEEE-RAS Int. Conf. Humanoid Robots*, Madrid, Spain, Nov. 2014, pp. 787–793, doi: 10.1109/HUMANOIDS.2014.7041453.

[13] G. Cannata and M. Maggiali, "Design of a tactile sensor for robot hands," in *Sensors: Focus on Tactile Force and Stress Sensors*. London, U.K.: IntechOpen, 2008. https://www.intechopen.com/chapters/6135, doi: 10.5772/6626.

[14] H. Nakamoto, F. Kobayashi, N. Imamura, and H. Shirasawa, "Universal robot hand equipped with tactile and joint torque sensors (development and experiments on stiffness control and object recognition)," in *Proc. 10th World Multi-Conf. Systemics Cybern. Inform.*, 2006, pp. 347–352.

[15] D. Goger, N. Gorges, and H. Worn, "Tactile sensing for an anthropomorphic robotic hand: Hardware and signal processing," in *Proc. IEEE Int. Conf. Robot. Autom.*, Kobe, Japan, May 2009, pp. 895–901, doi: 10.1109/ROBOT.2009.5152650.

[16] A. Schmitz, M. Maggiali, L. Natale, B. Bonino, and G. Metta, "A tactile sensor for the fingertips of the humanoid robot iCub," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2010, pp. 2212–2217, doi: 10.1109/IROS.2010.5648838.

[17] A. Schmitz, M. Maggiali, L. Natale, and G. Metta, "Touch sensors for humanoid hands," in *Proc. 19th Int. Symp. Robot Human Interact. Commun.*, Viareggio, Italy, Sep. 2010, pp. 691–697, doi: 10.1109/ROMAN.2010.5598609.

[18] A. Drimus, G. Kootstra, A. Bilberg, and D. Kragic, "Classification of rigid and deformable objects using a novel tactile sensor," in *Proc. 15th Int. Conf. Adv. Robot. (ICAR)*, Tallinn, Estonia, Jun. 2011, pp. 427–434, doi: 10.1109/ICAR.2011.6088622.

[19] A. Schmitz, P. Maiolino, M. Maggiali, L. Natale, G. Cannata, and G. Metta, "Methods and technologies for the implementation of large-scale robot tactile sensors," *IEEE Trans. Robot.*, vol. 27, no. 3, pp. 389–400, Jun. 2011, doi: 10.1109/TRO.2011.2132930.

[20] J. Schill, J. Laaksonen, M. Przybylski, V. Kyrki, T. Asfour, and R. Dillmann, "Learning continuous grasp stability for a humanoid robot hand based on tactile sensing," in *Proc. 4th IEEE RAS EMBS Int. Conf. Biomed. Robot. Biomechatronics (BioRob)*, Rome, Italy, Jun. 2012, pp. 1901–1906, doi: 10.1109/BioRob.2012.6290749.

[21] S. E. Navarro, N. Gorges, H. Worn, J. Schill, T. Asfour, and R. Dillmann, "Haptic object recognition for multi-fingered robot hands," in *Proc. IEEE Haptics Symp. (HAPTICS)*, Vancouver, BC, Canada, Mar. 2012, pp. 497–502, doi: 10.1109/HAPTIC.2012.6183837.

[22] H. Liu, J. Greco, X. Song, J. Bimbo, L. Seneviratne, and K. Althoefer, "Tactile image based contact shape recognition using neural network," in *Proc. IEEE Int. Conf. Multisensor Fusion Integr. Intell. Syst. (MFI)*, Hamburg, Germany, Sep. 2012, pp. 138–143, doi: 10.1109/MFI.2012.6343036.

[23] A. Schmitz, Y. Bansho, K. Noda, H. Iwata, T. Ogata, and S. Sugano, "Tactile object recognition using deep learning and dropout," in *Proc. IEEE-RAS Int. Conf. Humanoid Robots*, Madrid, Spain, Nov. 2014, pp. 1044–1050, doi: 10.1109/HUMANOIDS.2014.7041493.

[24] A.-M. Cretu, T. E. A. De Oliveira, V. P. Da Fonseca, B. Tawbe, E. M. Petriu, and V. Z. Groza, "Computational intelligence and mechatronics solutions for robotic tactile object recognition," in *Proc. IEEE 9th Int. Symp. Intell. Signal Process. (WISP)*, Siena, Italy, May 2015, pp. 1–6, doi: 10.1109/WISP.2015.7139165.

[25] L. Jamone, L. Natale, G. Metta, and G. Sandini, "Highly sensitive soft tactile sensors for an anthropomorphic robotic hand," *IEEE Sensors J.*, vol. 15, no. 8, pp. 4226–4233, Aug. 2015, doi: 10.1109/JSEN.2015.2417759.

[26] A. J. Spiers, M. V. Liarokapis, B. Calli, and A. M. Dollar, "Single-grasp object classification and feature extraction with simple robot hands and tactile sensors," *IEEE Trans. Haptics*, vol. 9, no. 2, pp. 207–220, Apr. 2016, doi: 10.1109/TOH.2016.2521378.

[27] M. Kaboli, R. Walker, and G. Cheng, "Re-using prior tactile experience by robotic hands to discriminate in-hand objects via texture properties," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Stockholm, Sweden, May 2016, pp. 2242–2247, doi: 10.1109/ICRA.2016.7487372.

[28] J. Qin, H. Liu, G. Zhang, J. Che, and F. Sun, "Grasp stability prediction using tactile information," in *Proc. 2nd Int. Conf. Adv. Robot. Mechatronics (ICARM)*, Hefei Tai'an, China, Aug. 2017, pp. 498–503, doi: 10.1109/ICARM.2017.8273213.

[29] M. Kaboli and G. Cheng, "Robust tactile descriptors for discriminating objects from textural properties via artificial robotic skin," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 985–1003, Aug. 2018, doi: 10.1109/TRO.2018.2830364.

[30] E. Torres-Jara and L. Natale, "Sensitive manipulation: Manipulation through tactile feedback," *Int. J. Humanoid Robot.*, vol. 15, no. 1, Feb. 2018, Art. no. 1850012, doi: 10.1142/S0219843618500123.

[31] S. Funabashi, G. Yan, A. Geier, A. Schmitz, T. Ogata, and S. Sugano, "Morphology-specific convolutional neural networks for tactile object recognition with a multi-fingered hand," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, Montreal, QC, Canada, May 2019, pp. 57–63, doi: 10.1109/ICRA.2019.8793901.

[32] G. Rouhafzay, A. M. Cretu, and P. Payeur, "Transfer of learning from vision to touch: A hybrid deep convolutional neural network for visuo-tactile 3D object recognition," *Sensors*, vol. 21, no. 1, pp. 1–15, Dec. 2020, doi: 10.3390/s21010113.

[33] Z. Kappassov, J.-A. Corrales, and V. Perdereau, "Tactile sensing in dexterous robot hands," *Robot. Auto. Syst.*, vol. 74, pp. 195–220, Dec. 2015, doi: 10.1016/j.robot.2015.07.015.

[34] A. Schneider, J. Sturm, C. Stachniss, M. Reisert, H. Burkhardt, and W. Burgard, "Object identification with tactile sensors using bag-of-features," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, St. Louis, MO, USA, Oct. 2009, pp. 243–248, doi: 10.1109/IROS.2009.5354648.

[35] Z. Pezzementi, E. Plaku, C. Reyda, and G. D. Hager, "Tactile-object recognition from appearance information," *IEEE Trans. Robot.*, vol. 27, no. 3, pp. 473–487, Jun. 2011, doi: 10.1109/TRO.2011.2125350.

[36] T. Bhattacharjee, J. M. Rehg, and C. C. Kemp, "Haptic classification and recognition of objects using a tactile sensing forearm," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Vilamoura-Algarve, Portugal, Oct. 2012, pp. 4090–4097, doi: 10.1109/IROS.2012.6386142.

[37] S. Dattaa, A. Khasnobishb, A. Konara, D. N. Tibarewalab, and R. Janarthananc, "Performance analysis of object shape classification and matching from tactile images using wavelet energy features," in *Proc. Int. Conf. Comput. Intell., Modeling Techn. Appl. (CIMTA)*, Kalyani, Kolkata, India, 2013, pp. 805–812, doi: 10.1016/j.protcy.2013.12.425.

[38] S. Luo, W. Mou, K. Althoefer, and H. Liu, "Novel tactile-SIFT descriptor for object shape recognition," *IEEE Sensors J.*, vol. 15, no. 9, pp. 5001–5009, Sep. 2015, doi: 10.1109/JSEN.2015.2432127.

[39] Y.-H. Liu, Y.-T. Hsiao, W.-T. Cheng, Y.-C. Liu, and J.-Y. Su, "Low-resolution tactile image recognition for automated robotic assembly using kernel PCA-based feature fusion and multiple kernel learning-based support vector machine," *Math. Problems Eng.*, vol. 2014, pp. 1–11, Jan. 2014, doi: 10.1155/2014/497275.

[40] J. M. Gandarias, J. M. Gomez-de-Gabriel, and A. Garcia-Cerezo, "Human and object recognition with a high-resolution tactile sensor," in *Proc. IEEE SENSORS*, Glasgow, U.K., Oct. 2017, pp. 1–3, doi: 10.1109/ICSENS.2017.8234203.

[41] J. M. Gandarias, A. J. Garcia-Cerezo, and J. M. Gomez-De-Gabriel, "CNN-based methods for object recognition with high-resolution tactile sensors," *IEEE Sensors J.*, vol. 19, no. 16, pp. 6872–6882, Aug. 2019, doi: 10.1109/JSEN.2019.2912968.

[42] S. Luo, W. Mou, K. Althoefer, and H. Liu, "Iterative closest labeled point for tactile object shape recognition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Daejeon, Korea (South), Oct. 2016, pp. 3137–3142, doi: 10.1109/IROS.2016.7759485.

[43] A. Albini, S. Denei, and G. Cannata, "Human hand recognition from robotic skin measurements in human–robot physical interactions," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Vancouver, BC, Canada, Sep. 2017, pp. 4348–4353, doi: 10.1109/IROS.2017.8206300.

[44] S. Luo, "Object perception through tactile images," Ph.D. dissertation, Dept. Informat., King's College, London, U.K., 2016, doi: 10.13140/RG.2.2.10040.44805.

[45] S. Pohtongkam and J. Srinonchat, "Object recognition using glove tactile sensor," in *Proc. Int. Electr. Eng. Congr. (iEECON)*, Khon Kaen, Thailand, Mar. 2022, pp. 1–4, doi: 10.1109/iEECON53204.2022.9741672.

[46] Z. Liao, W. Zhang, D. Kim, D. Im, K. Lim, and T. Miyoshi, "Recognition of the three-dimensional shape of objects grasped for PESA multi-fingered robot hand," in *Proc. IEEE Int. Conf. Real-Time Comput. Robot. (RCAR)*, Angkor Wat, Cambodia, Jun. 2016, pp. 472–476, doi: 10.1109/RCAR.2016.7784075.

[47] K. Weiss and H. Worn, "The working principle of resistive tactile sensor cells," in *Proc. IEEE Int. Conf.*, Niagara Falls, ON, Canada, Jul. 2005, pp. 471–476, doi: 10.1109/ICMA.2005.1626593.

[48] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

[49] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Lake Tahoe, NV, USA, 2012, pp. 1097–1105, doi: 10.1145/3065386.

[50] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[51] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[52] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.

[53] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-V4, inception-ResNet and the impact of residual connections on learning," 2016, *arXiv:1602.07261*.

[54] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1800–1807, doi: 10.1109/CVPR.2017.195.

[55] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 8697–8710, doi: 10.1109/CVPR.2018.00907.

[56] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

SOMCHAI POHTONGKAM was born in Nakhon Si Thammarat, Thailand, in 1980. He received the B.Eng., M.Eng., and D.Eng. degrees in electrical engineering from the Rajamangala University of Technology Thanyaburi, in 2002, 2010, and 2022, respectively.

He is currently a nuclear engineer for Canberra brand products in Thailand. He is also licensed as a radiation safety officer and serves to inspect radiation machines and nuclear material. His research interests include the development of robot sense, including electronic skin, nose, tongue, machine vision, and natural language processing. He specializes in nuclear electronics.

JAKKREE SRINONCHAT (Member, IEEE) received the B.Eng. degree in electrical engineering from the Rajamangala University of Technology Thanyaburi (RMUTT), in 1995, and the Ph.D. degree in electrical engineering from the University of Northumbria, Newcastle, U.K., in 2005. Since then, he has been a Researcher and a Lecturer with the Signal Processing Research and Laboratory, RMUTT. His research interests include speech and image processing, robot controller, machine learning, and applied AI for signal processing.