3차원 물체의 인식 성능 향상을 위한 감각 융합 신경망 시스템

論 文
54D-3-4

Neural Network Approach to Sensor Fusion System for Improving the Recognition Performance of 3D Objects

金 知 敬† · 董 聖 秀* · 李 鍾 浩**
(Ji Kyoung Kim · Sung Soo Dong · Chong Ho Lee)

Abstract – Human being recognizes the physical world by integrating a great variety of sensory inputs, the information acquired by their own action, and their knowledge of the world using hierarchically parallel-distributed mechanism. In this paper, authors propose the sensor fusion system that can recognize multiple 3D objects from 2D projection images and tactile informations. The proposed system focuses on improving recognition performance of 3D objects. Unlike the conventional object recognition system that uses image sensor alone, the proposed method uses tactual sensors in addition to visual sensor. Neural network is used to fuse the two sensory signals. Tactual signals are obtained from the reaction force of the pressure sensors at the fingertips when unknown objects are grasped by four-fingered robot hand. The experiment evaluates the recognition rate and the number of learning iterations of various objects. The merits of the proposed systems are not only the high performance of the learning ability but also the reliability of the system with tactual information for recognizing various objects even though the visual sensory signals get defects. The experimental results show that the proposed system can improve recognition rate and reduce learning time. These results verify the effectiveness of the proposed sensor fusion system as recognition scheme for 3D objects.

Key Words : Intelligent System, Neural Network, Sensor Fusion, Object Recognition, Haptic Recognition

1. Introduction

Human being recognizes the physical world with highly versatile sensibility. They adapt to physical world by integrating the sensory information acquired by their own action and knowledge. The human brain has the hierarchically parallel-distributed mechanism that can handle an enormous of sensory information.

A great deal of research has been introduced in relation to object and pattern recognition [1-10]. This is the difficult subject in the filed of image processing. There are two main approaches in object recognition. One method is model-based recognition. This scheme uses the 3D object representation in an object-centered coordination system [11], [12]. This method recognizes object by matching an input image with the model image that is obtained from the 3D object representation. For example, the technique of Yamane S [12] recognize a 3D object from a monocular-centered projection image based on an object model that uses a model that parameter to represent shape variability. However, some problems of model-based recognition are to consider that the 3D object model must be built in advanced and search range is extremely large during execution.

Second method is the view-based recognition that uses an object representation based on various 2D projection images of the 3D object [13], [14]. This method recognizes an object by matching an input image with multiple 2D projection images that are provided in advance. The engine-space method of Murase [13] attempts to represent a 3D object by using a subspace that extends engine-vectors of covariance matrix from learning images of 2D projection images. One problem of view-based recognition is that the matching with leaning images must be flexibly performed so that leaning images can correspond to the various 2D projection images of the 3D object.

Also, the Neocognitron has been used to design the parallel-structure of the vision system [15], [16]. The Neocognitron is the model that detects local features such as corner points or end-point from edge information of input image. The amount of the calculation required for recognition is too large because the principal aim of the Neocognitron is to design bio-mimic vision system. As a result, it is not usually used for view-based recognition of 3D object. But the Neocognitron has been mainly applied to character and face recognition because its recognition performance is very high for 2D object.

Additionally, It has been introduced that robotic gripper can identify various objects [17], [18]. Quian and Quio [17] suggested that their method are addressed in that the new type of gripper with the finger configuration of four circles are used instead of two parallel lines for recognizing objects as in [17]. This method uses thing that the key parameter of a symmetric four pin gripper is the distance of two pin centers on each finger. This depends on the

† 교신저자, 學生會員 : 仁荷大 工大 情報通信大學院 碩士
E-mail : jikyung_99@yahoo.co,kr
* 正 會 員 : 龍仁松潭大 디지털電子情報科 敎授 · 工碩
** 正 會 員 : 仁荷大 工大 情報通信工學部 敎授 · 工博
接受日字 : 2004年 11月 24日
最終完了 : 2005年 2月 15日

object shape and the closure property. The result is very successful and their idea is very fresh. But the experimental equipment in robotic gripper is very expensive and calculation is very complicated.

In this paper we propose the sensor fusion system for improving 3D object recognition rate and decreasing the number of learning iteration. Tactual information is obtained from reaction forces measured by grasping unknown object with four-fingered robot hand including pressure sensors at its fingertips. And visual information is obtained from CCD camera. This sensory information is fed to neural network for learning of each object.

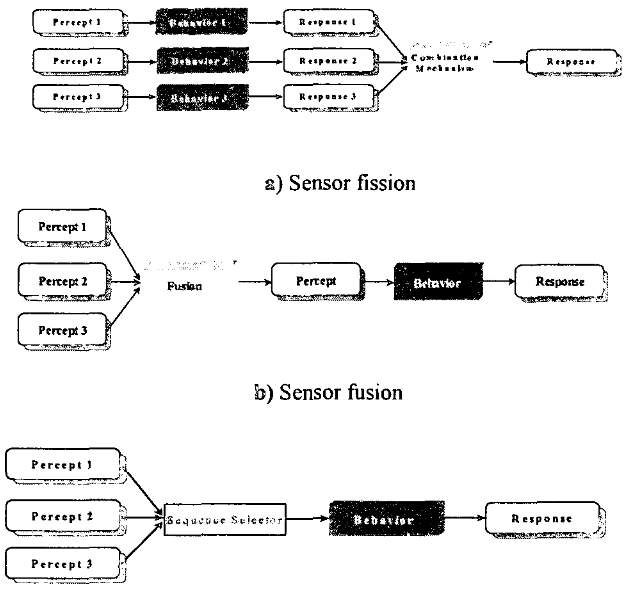
This paper is organized as follows. Section 2 presents the concept of sensor fusion. In section 3, intelligently optimized recognition system to classify various objects is proposed. The design procedure of proposed system is described in detail. Simulation and experimental results are provided to demonstrate learning ability and robustness of the system in section 4. Conclusions are drawn in section 5. The merits of the proposed systems are not only the high performance of the learning ability but also the reliability of the system with tactual information for recognizing various objects even though visual information has a defect.

2. The Paradigms of Sensor Fusion

Cognitive psychologists are studying sensory integration and inter-sensory perception in order to generate an accurate model of perception while engineers, computer scientists, and robotic researchers are building robots, which requires the mechanism, not theoretical models, for performing sensor fusion. Single sensor systems have not been completely successful for demanding tasks in navigation, target or goal recognition and general scene interpretation. The primary disadvantage of a single sensor system is its inability to reduce uncertainty. Uncertainty occurs when features are missing or when the sensor cannot measure all the relevant attributes of a percept and the observation is ambiguous. Sensor fusion also draws attentions for those who work in the Artificial Intelligence (AI) and Artificial Neural Network (ANN) areas. The issue is how to use information from one sensor to focus attention of others, and how to combine information from the multiple sensors to improve accuracy or confidence in recognition. Behavior-based systems can generally organize the perceptual information in three ways: sensor fission, action-oriented sensor fusion, and perceptual sequencing or sensor fashion as in [19], [20]. Fig.1 illustrates these concepts. Sensor fission is easily understood. For example, motor behavior requires the specific stimulus to produce a response thus each 'Percept' is directly connected to the 'Behavior' to control its outputs.

Action-oriented sensor fusion uses the construction of temporary representations that can be lead to 'Behavior'. Increased robustness is achieved by restricting the final percept to the requirements of a particular requirements of 'Behavior' and contexts as well as retaining the advantage of reactive control while more than one sensor to provide.

Fixed-action patterns sometimes require various stimulus to support their operation over time and space. Different sensors or different views of an environment may modulate a behavioral response when it unfolds. Perceptual sequencing allows the



a) Sensor fission

b) Sensor fusion

c) Sensor fashion

Fig. 1. Three paradigms of sensor fusion [19]

3. System Configuration

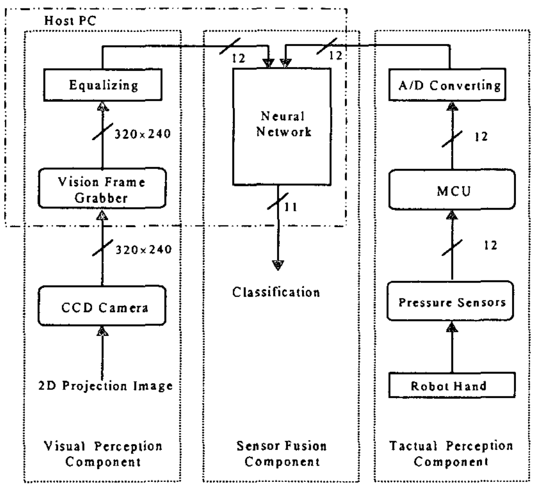


Fig. 2. The data flow of sensor fusion system

To implement recognition system, authors adopt Fig.1.b). But it is necessary to redesign each function block for effective recognition. So 'Behavior' and 'Response' block at the right side in Fig.1.b) is removed.

Fig. 2 shows the data flow of the proposed system. It has three components: visual perception component that consists of CCD camera and vision frame grabber, tactual perception component that consists of robotic hand and tactile sensor in it and sensor fusion component that uses neural network (NN). Fig. 3 illustrates the system configuration on experimental work space.

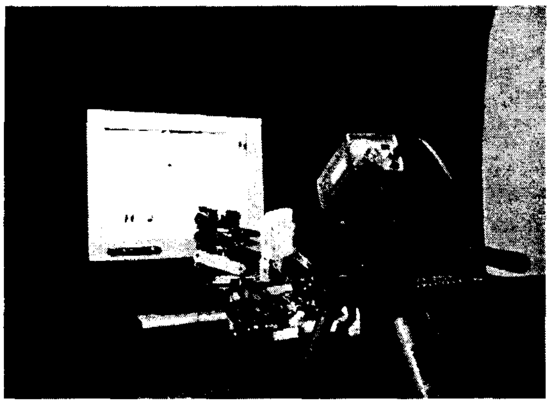


Fig. 3. The system configuration in experiment

3.1. Visual Perception Component

Image information from CCD camera is transmitted to personal computer (PC) by using vision frame grabber. Size of the image information in vision frame grabbers is 320×240 pixels and this information is transmitted with format of 8 bit gray scale. In our experiments, size of this information is too large for use. So this image information is transformed. Transformed size is given by (1) and (2)

$$20\times 20[\text{pixel}] = 1[\text{PU}] \tag{1}$$

$$i(k,l) = \sum_{x=20k}^{20k+19}\left(\sum_{y=20l}^{20l+19} v(x,y)\right)$$
$$0 \leqq k \leqq 15, \quad 0 \leqq l \leqq 11$$

$$v(x,y) = \begin{cases} 1, & pixel\ intensity \geq 123 \\ 0, & otherwise \end{cases}$$
$$0 \leqq x \leqq 319, 0 \leqq y \leqq 239 \tag{2}$$

where x denotes the number of rows, y is the number of columns in 320×240[pixel]. k denotes the number of rows and l is the number of columns in 16×12[PU].

Therefore transformed size is 16×12[PU]. Criterion of this transformation is chosen by experiment for minimal control. Simultaneously this signal is digitalized. That is, because $v(x, y)$ has 0 through 255 intensity, it is transformed into 0 or 1 value according to the threshold value which is determined as 123 through experiment. So, $i(k, l)$ has 0 through 400 values. But we have to use binary code. So this value is finally changed to the binary signal by (3).

$$i_n = \begin{cases} 0, & i(k,l) < 150 \\ 1, & i(k,l) \geq 150 \end{cases} \tag{3}$$

50 that is the multiplication of 3/8 and 400 of $i(k, l)$ is used for threshold value. And this threshold selects binary code. Finally transformed binary signal is used for input of NN. The number of NN inputs for visual recognition is 12. Transformation process of image information is illustrated in Fig. 4. Color of object is white and color of background is black in experimental work space. Single CCD camera captures various 2D projection images of 3D objects based on view-based recognition. So the 16×12 matrix is used for 2D projection image. Each row is used for NN input. For example, $i_o \sim i_{15}$ is used for the first input neuron and $i_{175} \sim i_{191}$ is used for the last input neuron. So the total number of neurons are 12.

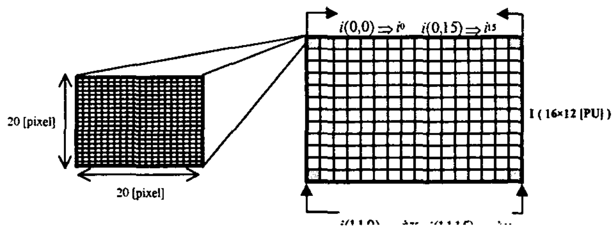


Fig. 4. The transformation of image information

3.2. Tactual Perception Component

For haptic recognition, tactual perception component consists of twelve pressure sensors in four-fingered gripper and AVR micro-control unit (MCU). Tactile sensors can treat the value of the pressure force that have sensing range of 0 through 10 lbs. For receiving tactile information of the pressure sensors, the internal cache unit in MCU is used. It receives analog sensory inputs and temporarily stores in the cache unit and then converts analog signal into digital signal. After conversion, it transfers digitalized pressure information to NN. Also the role of MCU is the control of robotic arm.

The approach used in this study is that one of the four fingers of the robotic hand provides contacts of position guiding for the grasped object and the other three fingers measure the reaction forces at their each contact points with grasping object. Finger A has three pressure sensor, and the other three fingers also have three pressure sensors per each finger. So twelve pressure force intensity is used for NN input. One important assumption is that there is no tangential characteristic of friction at the contact point. Pressure sensors of each finger and grasp contact model are illustrated in

Fig. 5.

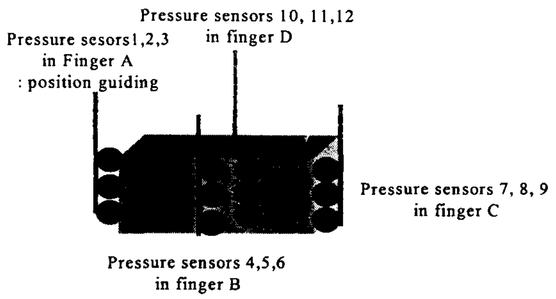


Fig. 5. The scheme of grasping contacts

When robotic hand grasps object, it is allowed to impact the object in any angular position. We analyze force patterns that are measured by grasping objects in different orientations in the workspace of fixed object position. And it is shown that geometry of each object yields different characteristic with its contact force patterns. If w is the length of one side of the object, finger A is firstly located on $w/3$ or $w/5$ of the one side. That is, the first position for training set is $w/3$ and the first position for test set is $w/5$. Once the finger A is placed on these positions, the other fingers B, C, D are automatically relocated because of fixed inter-contact distance. Then finger A moves to next position which is $2w/3$ or $2w/5$. Incremental value 1/3 or 1/5 is applied to all sides of all objects for making force patterns.

One more assumption is that the objects are not allowed to deform by excessive force. Consequently, patterns of contact force are generated by the finite element method for differently rotating contact positions on object and are used to define characteristics of each object shape and mass. These patterns form the training sets and test sets that are fed to sensor fusion component. And they are used to train and test the shape recognition. Fig. 6 presents some examples of robotic grasp positions along the object side and Fig. 7 illustrates the shapes of eleven objects.

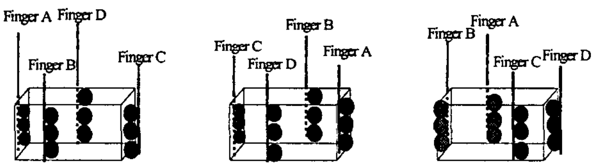


Fig. 6. Some examples of robotic grasp positions

Object A is triangular shape, and objects B, C, D, E, G are square shape. Object F is pentagonal shape and object H is circle shape with a hole and objects I, J are cylinder shape. Finally, object K is ball shape. Particularly objects B, C, D, E, G or objects I, J are very similar in shape and size. But the noticeable difference among these objects is mass.

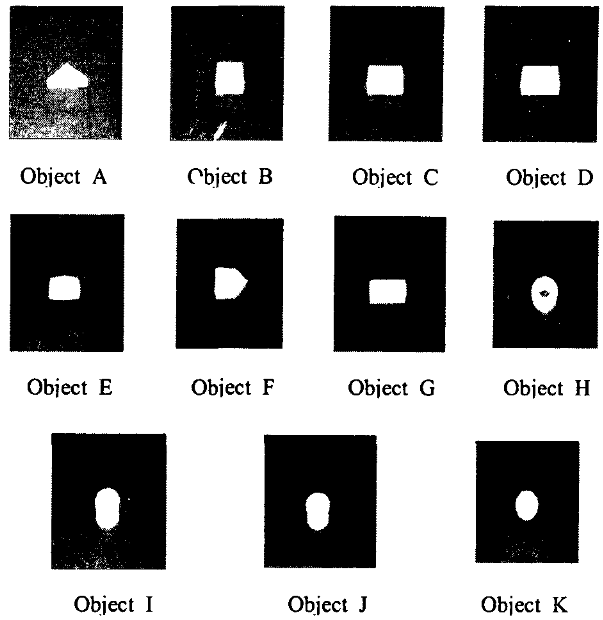


Fig. 7. The shapes of eleven objects

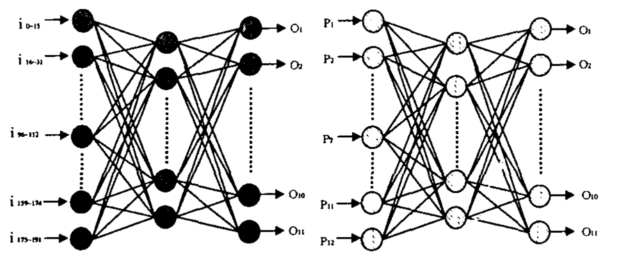
3.3. Sensor Fusion Component

The purpose of the proposed sensor fusion component is to improve performance of 3D object recognition rate and reduce the number of learning iterations. Because visual sensor coordinates with tactile sensors, this fusion component is faster than the recognition system using single vision information. And this component has high performance on 3D object recognition.

Additionally, this component is also a great of value in that it has reliability of recognition system even thought vision information is out of order. In other words, this fusion component consists of vision and tactile information as input of NN. But when visual perception component has insufficient information or malfunction, NN does a function as recognition system with a few tactile sensors. So, visual input signal and tactile input signal of NN is superimposed. Vision input signal is connected with vision frame grabber and its input size for NN is 12. And tactile input signal for NN is 12. So the total number of input neurons are 24 for sensor fusion NN. Pressure sensors in finger A, B, C and D are used for input signal of NN. Finger A also has a role as guide of contact position when robotic hand grasps objects.

Fig. 8 illustrates each NN structure of image and tactile information and Fig. 9 represents the superimposed NN structure for sensor fusion process.

This component for recognizing objects follows multi-layered NN in the form of a NN structure as in [21]. Specification used in this component is illustrated in Table 1. Data and method of learning are explained in section 4.



[NN for tactile information] [NN for visual information]

Fig. 8. The NN structures for processing image and tactile information.

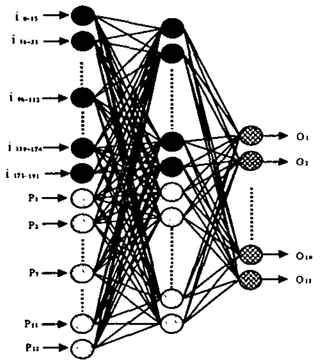


Fig. 9. The NN structure for sensor fusion

Table 1. The NN specification for sensor fusion component.

The number of inputs	Image(12) + Pressure(12)
The number of hidden neuron	Image(10) +pressure(10)
The number of outputs	11
Learning rate	0.45
Tolerance	0.15
Max iteration	2000
Learning method	Back propagation algorithm

The factor affecting the number of hidden layers in the NN is the size of the training set. Since our training set with maximum 33 data set is a small, the optimal number of hidden layers is estimated to be from 1 to 3. Nevertheless, training experiment is conducted with from 1 to 5 hidden layers with different numbers of neurons in each hidden layer. One hidden layer configuration with 20 hidden neurons is found to give the best results in terms of lower classification error. The best learning rate is found to be 0.45 for fast convergence and initial/ending tolerances of 0.15 give the best estimation percentage.

4. Experimental Method and Results

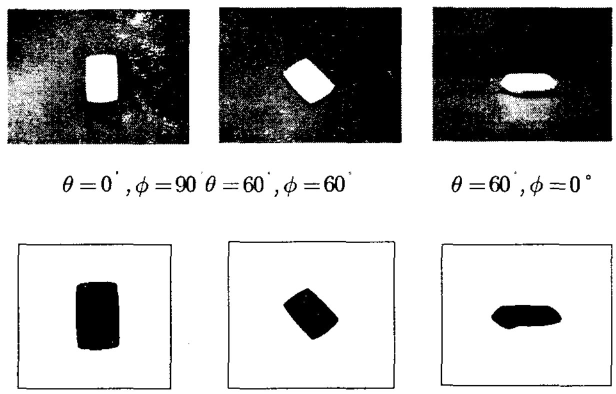
We perform experiments to evaluate the basic efficiency of the proposed system for proper recognition rate and learning time. Experimental results compare recognition rate and the number of learning iterations using fusion method with that of visual information.

4.1. The Training of Sensor Fusion Data

It has been said that the vision system of a living organism performs recognition of object by mentally rotating or transforming the target object [22].

According to the fact as in [22], we use images that are photographed by rotating 30 or 60 degrees per each object along vertical axis or horizontal axis.

Fig. 10 shows some examples of image postures for learning. This is object E. Some part of leaning process is presented. The object to be recognized is placed within the field of view of the camera. It is assumed that the feature projection image of the object can be observed from any view point. Training set of image information is as follow. The spherical observation surface is divided into both θ and ϕ according to the mesh that is equally spaced with latitudinal and longitudinal lines per 60 degrees. θ is the yaw-angle of range $0^\circ \sim 360^\circ$ and ϕ is the pitch-angle of range $0^\circ \sim 180^\circ$ in hemisphere. This data set consists of 2D projection images at the intersections (6×3=18). Test set of image information is as follow. The spherical observation surface is divided into both θ and ϕ according to the mesh that is equally spaced with latitudinal and longitudinal lines per 30 degrees. This test set consists of 2D projection images placed at the intersections (12×6=72). The position of a view point is defined as the pose of the object with respect to the camera.



$\theta=0^\circ, \phi=90$ $\theta=60^\circ, \phi=60^\circ$ $\theta=60^\circ, \phi=0^\circ$

Fig. 10. Three postures examples of object E and their 2D images for learning.

For illustrative purpose, each finger is assigned with the characters from A to D. Finger A has three pressure sensors and the others also have three pressure sensors per each finger named from P1 to P12.

Object F gives the most complicated patterns because it has five sides. Shape for object C, D, E, G is similar but the mass for them is different. So each object has different force values to prevent slipping in gripper. Because cylinder-shaped objects I and J have the same reason, so they can only be discriminated by tactual sensing.

Ball-shaped object K shows unique patterns due to its symmetry. When gripper grasps object K, all the grasp position generates same reaction force.

For making training data and test data, finger A is located on $w/3$, $2w/3$, $3w/3$ and $w/5$, $2w/5$, $3w/5$, $4w/5$, $5w/5$ of all sides for all objects. Once the finger A is placed on these position, the other finger B, C, D are automatically relocated because of fixed inter-contact distance. So, object A has 9 training patterns and 15 test patterns. Object F has 15 training patterns and 25 test patterns since it has five side. Object B, C, D, E, G, H, I, J, K have 12 train patterns and 20 test patterns. Table 2 shows some examples of classification with values of twelve pressure sensors for all objects.

Table 2. Some examples of classification by test set using pressure sensors

The values of Pressure Sensors												
Finger A			Finger B			Finger C			Finger D			Object
P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	
4.48	4.50	4.49	3.21	3.23	3.24	4.49	4.51	4.50	3.20	3.23	3.24	A
4.39	4.40	4.40	3.39	3.41	3.41	4.38	4.41	4.41	3.39	3.40	3.41	A
2.24	2.25	2.25	2.29	2.31	2.31	2.21	2.19	2.19	2.30	2.30	2.31	B
2.13	2.16	2.17	2.15	2.14	2.14	2.12	2.13	2.13	2.15	2.13	2.13	B
2.49	2.50	2.49	4.10	4.12	4.13	2.48	2.47	2.48	4.11	4.11	4.12	C
2.62	2.63	2.62	5.13	5.13	5.16	2.60	2.63	2.64	5.12	5.12	5.15	C
3.76	3.77	3.77	5.54	5.52	5.52	3.80	3.80	3.81	5.55	5.53	5.54	D
3.72	3.71	3.78	5.49	5.50	5.51	3.82	3.83	3.83	5.50	5.50	5.51	D
5.53	5.53	5.52	6.21	6.20	6.21	5.48	5.49	5.48	6.20	6.21	6.20	E
5.63	5.66	5.65	6.15	6.17	6.18	5.66	5.68	5.68	6.16	6.17	6.16	E
1.92	1.93	1.92	3.34	3.35	3.34	2.00	2.04	2.04	3.35	3.36	3.35	F
1.96	1.99	1.98	3.29	3.31	3.31	1.96	1.97	1.96	3.29	3.30	3.30	F
7.71	7.70	7.70	8.02	8.05	8.04	7.69	7.71	7.70	8.00	8.02	8.04	G
7.64	7.67	7.66	8.11	8.15	8.14	7.64	7.62	7.63	8.10	8.13	8.14	G
5.55	5.56	5.55	5.55	5.56	5.52	5.55	5.56	5.54	5.55	5.55	5.56	H
5.59	5.61	5.60	5.59	5.61	5.60	5.59	5.61	5.62	5.60	5.58	5.69	H
1.02	1.06	1.05	1.56	1.58	1.58	1.04	1.07	1.07	1.55	1.57	1.57	I
4.96	4.95	4.96	5.01	5.02	5.01	4.95	4.97	4.97	5.01	5.02	5.01	J
6.24	6.25	6.24	6.24	6.25	6.24	6.25	6.24	6.25	6.24	6.24	6.25	K

When comparing the contact force patterns of the eleven objects, it is found that object G has the heaviest values and the next values is object K. Object I has the least values. Ball-shaped object K has a discriminative contact force patterns with clear. That is, this object has similar pressure values for each finger due to its symmetry that all the grasping position generates. Regular square-shaped object B also has unique force patterns with the same reason. Above all, it is important that similar-shaped objects B, C, D, E, G or I, J are classified by contact force using pressure sensors. These objects have different contact force patterns since they have a different mass.

4.2. Results

We evaluate performance by the following criteria.

Firstly, how much accurately does the proposed sensor fusion system identify objects in comparison with single vision information? Secondly, how many does this sensor fusion system reduce the number of learning iterations in comparison with single vision information? Thirdly, it is demonstrated that the haptic recognition system has better performances better than vision-based recognition for similar-shaped object.

Fig. 11 shows the recognition rate for eleven objects. The figure indicates that recognition of sensor fusion is better than that of vision sensing. This reason is that sensor fusion system has more sufficient feature extraction than single haptic or visual recognition system. And haptic recognition performs better than visual recognition for similar shapes of objects. That is, classification results about objects B, C, D, E, G and objects I, J are better than that of visual recognition. But visual recognition is more effective than haptic recognition for triangular-shaped object A because this object has a unique shape in comparison with other objects. Also, ball-shaped object K and circle-shaped with a hole object H have a good visual recognition rate about 97.5% due to their unique shape. But sensor fusion and haptic recognition are also more effective than visual recognition for these two objects K, H. Recognitions using visual sensor for these seven objects are less than 73% but recognitions using tactile sensor are more than 85%.

And classification for all eleven objects by using sensor fusion is more than 93 %. This result verifies that the proposed sensor fusion system in this paper has the best efficiency for object recognition. This system is especially efficient for recognition of similar-shaped objects.

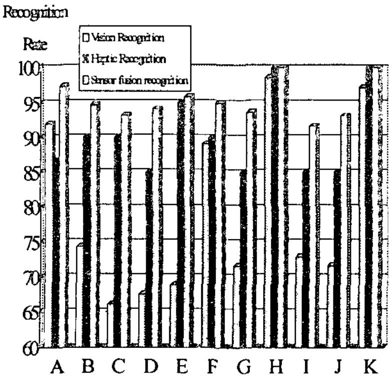


Fig. 11. The comparison of recognition rates.

Fig. 12 illustrates that the sensor fusion system reduces the number of learning iterations or improves the recognition rates for each object. This figure verifies that sensor fusion system can improve learning speed in comparison with the visual recognition using 2D projection image.

Vertical axis denotes recognition rate and horizontal axis is the number of learning iterations in Fig. 12. Comparing the number of learning iterations for eleven objects, it is found that all objects have

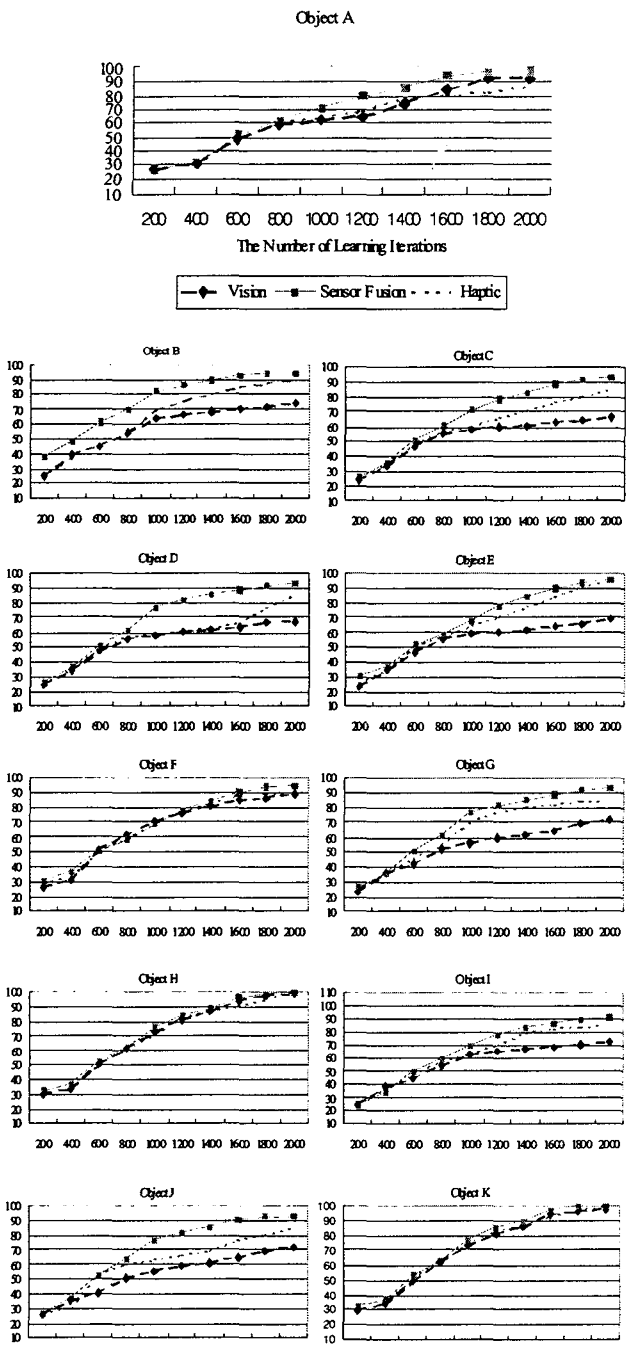


Fig. 12. The comparisons of recognition rates with the number of learning iterations.

the least learning iterations when they use sensor fusion recognition. In particular, when similar-shaped objects B, C, D, E, G or I, J are to be recognized visually, they make little increase of recognition rate with respect to increasing the learning iteration. But sensor fusion with haptic recognition for these seven objects can increase the recognition rate more effectively than visual recognition alone. Visual recognition for triangular-shaped object A tends to increase faster than haptic recognition above learning iterations of 1500 since this object has sufficient feature extraction with distinguished shape. And object H, K generally have similar slope of learning speed. That is, the slopes of learning speeds for fusion, haptic and visual recognition of objects H and K are very similar for these two objects. Sensor fusion NN system using additionally pressure sensors compensate for 2D projection image error.

4.3. Statistical Analysis

We analysis performance by the following criteria.

How much differently does the proposed sensor fusion system have the mean differences of object recognition rates in comparison with single vision information or haptic information?

For verifying the difference of the mean for vision *Vs* fusion and haptic *Vs* fusion recognition rate, T-test is used for recognition rate with learning iteration. Because sample size is less than 30 T-test is more useful than Z-test. Firstly we have to check whether the variance of comparison group is equal or not. If the variance is equal, equal-variance T-test is used for the mean difference. If the variance is unequal, unequal-variance T-test is used for the mean difference.

Method of statistical analysis is as follow.

1. Check the null and alternative hypothesis for the variance of two comparison group.

$$H_0 : \sigma_1^2 = \sigma_2^2 \ \ versus \ \ H_1 : \sigma_1^2 \neq \sigma_2^2 \tag{4}$$

Where σ^2 denotes variance. Null hypothesis is defined as H_0. Alternative hypothesis is defined as H_1.

If the P-value is less than 0.05 and F statistic value is much larger than 1, then you would say that you reject the null hypothesis at the 0.05 significance levels. That is, if the null hypothesis is rejected, you conclude that there is a difference of variance for comparison groups. So you have to use unequal variance T-test for the mean difference of comparison group.

If the P-value is greater than 0.05,and F statistic value is approximate to the 1 you would say that there is not enough evidence from the data to conclude that the null hypothesis is false. That is, you can conclude that there is not a difference of variance for tow comparison groups. So you have to use equal-variance T-test.

2. Then check the null and alternative hypotheses for the mean of comparison group.

$$H_0 : \mu_1 = \mu_2 \ \ versus \ \ H_1 : \mu_1 \neq \mu_2 \tag{5}$$

where μ denotes mean.

If the p-value corresponding to the T-statistic value is less than 0.05 you would say that you reject the null hypothesis at the 0.05 significance level. That is, if the null hypothesis is rejected, you conclude that there is enough difference of mean for tow groups.

If the p-value corresponding to the t-statistic value is larger than 0.05 you would say that you can't reject the null hypothesis at the 0.05 significance level. That is, if the null hypothesis is accepted, you conclude that there is not enough difference of mean for tow groups.

Table 3 illustrates the variables about F-test and T-test of vision *Vs* fusion and haptic fusion *Vs* recognition rate for 11objects when the number of learning iteration is 2000. Total number of the variable is 22. Variable is 11 for 11 vision object recognition rate and 11 for fusion recognition rate.

Table 3. Variables for F-test and T-est when learning iteration is maximum 2000

Object name	Variables (recognition rates) when learning iteration 2000		
	vision	fusion	haptic
Object A	91.89	97.29	86
Object B	74.32	94.66	90
Object C	66.21	93.24	85
Object D	67.56	94.05	85
Object E	68.91	95.94	95
Object F	89.1	95	90
Object G	71.62	93.64	85
Object H	98.6	100	100
Object I	72.9	91.68	85
Object J	71.6	93.1	85
Object K	97.2	100	100

Table 4 presents results of F-test and T-test of vision *Vs* fusion recognition rate according to the variables in the tables 3. F-statistic value is 20.49168 and p-value corresponding to the F-statistic value is 0.000023 in the case of finishing the learning. As the result of estimating the F-test, F-statistic value 20.49 is much larger than 1 and P-value corresponding to the F-statistic value 0.000023 is much less than 0.05 significance levels. These values are rejected for null hypothesis of variance. So unequal- variance T-test is selected. As the result of estimating unequal-variance T-test, T-statistic value is -4.2282 and P-value corresponding to the T-statistic value is 0.00141.These values are rejected for null hypothesis of mean. So you would say that there have enough evidence in mean difference for visual recognition rate *Vs* fusion recognition rate.

Table 5 presents results of F-test and T-test of haptic *Vs* fusion recognition rate according to the variables in the tables 3. F-statistic value is 4.82539 and P-value corresponding to the F-statistic value is 0.010178 in the case of finishing the learning. As the result of estimating the F-test, F-statistic value 4.82539 is much larger than 1 and P-value corresponding to the F-statistic value 0.010178 is much less than 0.05 significance levels. These values are rejected for null hypothesis of variance. So unequal-variance T-test is selected. As the result of estimating unequal-variance T-test, T- statistic value is -2.84515494 and p-value corresponding to the t-statistic value is 0.0129733. These values are rejected for null hypothesis of mean. So you would say that there have enough evidence in mean difference for visual recognition rate *Vs* fusion recognition rate.

Table 4. The result of F-test and T-est for vision Vs fusion recognition rate when learning iteration is maximum 2000

	vision	fusion
Mean	79.08273	95.32727
Variance	154.8089	7.554722
The number of observation	11	11
DOF	10	10
F-statistic value	20.49168	
P(F<=f) value	2.35E-05	
Difference of assumption mean	0	
T-statistic value	-4.22823859	
P(T<=t) value	0.00141661	

Table 5. The result of F-test and T-est for haptic Vs fusion recognition rate when learning iteration is maximum 2000

	hatic	fusion
Mean	89.63636	95.32727
Variance	36.45455	7.55472
The number of observation	11	11
DOF	10	10
F-statistic value	4.825399	
P(F<=f) value	0.010178	
Difference of assumption mean	0	
T-statistic value	-2.84515494	
P(T<=t) value	0.0129733	

Table 6. The result of T-est for vision Vs fusion and haptic Vs fusion in all learning stages.

Learning iteration	vision *Vs* fusion		haptic *Vs* fusion	
	T value	P value	T value	P value
200	-2.710	0.0134	-1.850	0.0790
400	-1.925	0.0684	-1.516	0.1449
600	-4.428	0.0002	-3.239	0.0041
800	-3.839	0.0010	-3.433	0.0026
1000	-4.690	0.0001	-3.928	0.0008
1200	-5.293	0.0001	-3.913	0.0015
1400	-5.253	0.0002	-3.903	0.0020
1600	-4.449	0.0009	-3.776	0.0020
1800	-4.236	0.0013	-3.986	0.0013
2000	-4.228	0.0014	-2.845	0.0012

From the table 4 and table 5 we conclude that recognition rate using sensor fusion system is more effective than single sensor system in the view of mean for all objects.

Table 6 illustrate the result of T-test according to the number of learning iteration. In all learning stages except learning stage 200 and 400, P-value corresponding to the T-statistic value is less than 0.05 significance level. we conclude that have enough evidence for mean difference between single and fusion sensor in all learning stages except learning stage 200 and 400.

Learning iteration 200 through 1000 uses equal-variance T-test and learning iteration 1200 through 2000 uses unequal-variance T-test.

5. Conclusion

We proposed sensor fusion system that consists of NN to achieve high performance on 3D recognition rate and reduction of learning iterations. The model also aims for tactile information to improve the system ability for recognizing 3D objects when visual information has defects. The proposed model includes three components called visual perception component, tactile perception component, and sensor fusion component.

We apply our model for recognition of eleven objects. Computer simulation of the proposed system proves that this system is useful for object recognition. The estimated performance indicates that this system can improve accuracy of recognition and speed of learning time with sensor fusion. Recognition using tactile sensors is better than that of single vision sensors for similar-shaped objects.

In our model, multi-layered NN in the sensor fusion component uses the coordination method of learning with visual senses and tactile senses. When a recognition system uses the proposed system, the system itself can autonomously acquire training data for the NN by driving its own actuators and sensing the physical world with its own pressure sensors.

It is anticipated that we will have to study how to recognize the colors of objects and apply to practical objects. And we should also study how to effectively cooperate the recognition systems with motor systems and how to exchange information among different hierarchical levels if the number of hierarchical sensor levels is increased.

Acknowledgement

This research was sponsored by Inha University (INHA-31544). We wish to thank them for their generous financial assistance.

References

[1] Tae-Woo Lee, Chang-Ik Jun, "A Study on the Pattern Recognition of EMG Signals for Head Motion Recognition", Trans. KIEE, Vol.49D, No.2, pp.103-110, JUL.2004.

[2] In-Seok Cho, Jin-Hyun Park, "A Study on Dynamic Hand Gesture Recognition Using Neural Networks", Trans. KIEE, Vol.53D, No.1, pp.22-31, JAN.2004.

[3] Young-Jae Ryoo, Eui-Sun Kim, "Recognition of Road Direction for Magnetic Sensor Based Autonomous Vehicle", Trans. KIEE, Vol.52D, No.9, pp.526-532, SEP.2003.

[4] Ho-Seon Choi, Won-Kyu Jang, "Snorer-Dependent Snore Recognition Using LPC Cepstral Coefficients", Trans. KIEE, Vol.52D, No.9, pp.554-559, SEP.2003.

[5] In-Mo Ahn, Kee-Sang Lee, "Recognition of Patterns and Marks on the Glass Panel of Computer Monitor", Trans. KIEE, Vol.52D, No.9, pp.35-41, SEP.2003.

[6] Young-No Kim, Jae-Chul Kim, "The Detection of Partial Discharge Signal by the Measurement of an Electromagnetic Wave and Pattern Recognition Technique", Trans. KIEE, Vol.51A, No.6, pp.276-283, 2002.

[7] Bongkyu Lee, Yohan Yang, "An Implementation of Generalized Second-Order Neural Networks for Pattern Recognition", Trans. KIEE, Vol.51D, No.10, pp.446-452, OCT.2002.

[8] Hyun-Chul Cho, Keeseong Lee, "The 3-D Underwater Object Recognition Using Neural Networks and Ultrasonic Sensor Fabricated with 1-3 Type Piezoelectric Composites", Trans. KIEE, Vol.50C, No.7, pp.324-328, JUL.2001.

[9] Sung-Soo Kim, "Pulse Code Signal Recognition using Integra-Normalizer", Trans. KIEE, Vol.49D, No.8, pp.491-494, AUG.2000.

[10] Chung-Soo Lim, Chong-Ho Lee, "Neural Network Design for Spatio-temporal Pattern Recognition", Trans. KIEE, Vol.49D, No.7, pp.1464-1471, JUL.1999.

[11] Y. Sato, S.Tamamura, "3D object recognition using parameterized model", Trans. Inf. Soc. Japan, vol. 31, pp. 56-67, 1990.

[12] S. Yamane, M. Izumi, K. Fukunaga, "Modelbased object position and orientation prediction", Trans. IEICE, vol. J79-D-Ⅱ, pp. 165-173, 1996.

[13] H. Murase, SK. Nayar, "3D object recognition from appearance", Trans. IEEE, vol. J77-D-Ⅱ, pp.2179-2187, 1994.

[14] T. Amano, S. Hiura, A. Yamaguchi, "Eigen space approach for pose detection with range images", Trans. IEICE, vol. J80-D-Ⅱ, pp. 1136-1143, 1997.

[15] AY. Cheng, DS. Yeung, "Sensitivity analysis of neocognitron", Trans. IEEE, Syst. Man. Cybern., vol. 29, pp. 238-249, August 1999.

[16] S. Sato, J. Kuroiwa, H. Aso, S. Miyake, "A rotation invariant neocognitron", Trans. IEICE, vol. J81-D-Ⅱ, pp.1365-1374, 1998.

[17] Wen-han Quain, Hong Quio, S.K. Tso, "Synthesizing two-fingered grippers for positioning and identifying objects", Trans. IEEE, Syst. Man. Cybern., vol. 31, pp. 602-614, 2001.

[18] Makoto Kaneko, Tatsuya Shirai, Toshio Tsuji, "Scale-Dependent grasp", Trans. IEEE, Syst. Man. Cybern., vol. 30, pp. 806-816, November 2000.

[19] R. Murphy, "Biological and cognitive foundations of intelligent sensor fusion", Trans. IEEE, Syst. Man. Cybern., vol. 26, pp. 42-51, January 1996.

[20] R. Murphy, "Dempster-shafter theory for sensor fusion in

autonomous mobile robot", Trans. IEEE, Robot. Automat., vol. 14, pp. 197-206, April 1999.
[21] Haykin, Neural Networks, Prentice Hall, 1999.
[22] DI. Perrett, MV. Ormam, E. Ashbridge, "An account of generalization of recognition with mental transformations", Cognition, vol. 67, pp.111-145, 1998.

저 자 소 개

김 지 경(金 知 敬)
He received the B.S. degrees in Electronic Engineering from Inha University in 2003 and the M.S. degree in information technology and telecommunication Engineering from Inha University in 2005.
Tel : 032-860-7396
Fax : 032-863-5822
E-mail : jikyoung_99@yahoo.co.kr

동 성 수(董 聖 秀)
He received the B.S. and M.S. degrees in Electrical Engineering from Inha University in 1990 and 1992. He is currently a Ph.D. candidate in Inha University. He is currently an Assistant Professor in the Department of Digital Electronics & Information at the Yongin- Songdam College.
Tel : 032-330-9292
Fax : 032-330-9290
E-mail : ssdong@ysc.ac.kr

이 종 호(李 鍾 浩)
He received the B.S. and M.S. degrees in Electrical Engineering from Seoul National University of Korea in 1976 and 1978, respectively. He received Ph.D. degree from Iowa State University in 1986. During 1979-1982, he was an instructor on electrical engineering department at the Korea Naval Academy. He was an assistant professor at the University of Notre Dame from 1986 to 1989, and then he moved to Inha University where he is a professor in information and communication engineering. In the year of 1994, he was visiting the Institute of Brain and Neural Systems at Brown University, USA. He is currently working on designing intelligent chips and directing the Super Intelligence Technology Center at Inha University. His research interests include neural network, evolvable hardware and intelligent system.
Tel : 032-860-7396
Fax : 032-863-5822
E-mail : chlee@inha.ac.kr