Visual–tactile object recognition of a soft gripper based on faster Region-based Convolutional Neural Network and machining learning algorithm

Chenlei Jiao, Binbin Lian, Zhe Wang, Yimin Song and Tao Sun

Abstract
Object recognition is a prerequisite to control a soft gripper successfully grasping an unknown object. Visual and tactile recognitions are two commonly used methods in a grasping system. Visual recognition is limited if the size and weight of the objects are involved, whereas the efficiency of tactile recognition is a problem. A visual–tactile recognition method is proposed to overcome the disadvantages of both methods in this article. The design and fabrication of the soft gripper considering the visual and tactile sensors are implemented, where the Kinect v2 is adopted for visual information, bending and pressure sensors are embedded to the soft fingers for tactile information. The proposed method is divided into three steps: initial recognition by vision, detail recognition by touch, and a data fusion decision making. Experiments show that the visual–tactile recognition has the best results. The average recognition accuracy of the daily objects by the proposed method is also the highest. The feasibility of the visual–tactile recognition is verified.

Keywords
Soft gripper, visual–tactile recognition, faster RCNN, machine learning algorithm

Introduction
Soft grippers made of soft materials have caused widespread concern for its capability of holding objects with various shapes, interacting effectively with unstructured environments, and performing tasks in a more dynamic manner. Until now, there have been a large variety of soft grippers, including those who are made from elastomeric pneumatic actuators, shape memory alloy (SMA)-driven, shape memory polymer, dielectric elastomer, ionic polymer–metal composites, or electro-adhesive polymer. Among these, soft and smart materials were mainly applied and developed. Besides investigating the new materials, some research proposed a novel technique for direct 3D printing of soft pneumatic actuators and fully multimaterial three-dimensional (3D) printed soft gripper. With the innovations in material development, structural design, and manufacturing, the soft gripper has been utilized in building integrated systems for different application scenarios, such as rehabilitation and assistance. From the existing research on soft grippers, it is found that the majority are on the material and manufacturing techniques, the research on the practical application...
of the soft gripper are still deficient, which is, in fact, an important issue in constructing autonomous system regarding the soft gripper as the execution unit.

Object recognition is the first problem to be solved in the application of soft gripper. In general, visual and tactile recognitions are the fundamental approaches that are commonly adopted in the research on grippers. Visual recognition uses a camera to obtain the object image and identify the features.\textsuperscript{20,21} However, there are several factors that affect the performance of extracted features and limit the performance of vision-based methods, such as scaling, rotation, translation, and illumination.\textsuperscript{20,22} Moreover, some characteristics of the object, for instance, hardness, temperature, and weight, cannot be identified by vision. Tactile recognition receives information from the tactile sensors installed in the grippers. For example, bio-Tac sensors are adopted to obtain vibration of the object or texture of the surface to classify object material and shape.\textsuperscript{23–25} Angle and pressure sensors are utilized to analyze the bending information of the finger for recognizing different objects.\textsuperscript{26} In summary, the influence of scaling can be removed in the tactile recognition as the real dimension and shape of the interacted object are mapped to the tactile sensor directly. In addition, tactile recognition can be used to capture properties like texture, roughness, spatial features, compliance, and friction,\textsuperscript{27,28} which are difficult to be recognized by vision. Hence, it seems to be promising to adopt the tactile recognition in the autonomous system containing soft gripper. Since the existing tactile recognition methods are applied mainly to rigid grippers, there are two main questions if we want to apply them to the soft gripper: (1) Due to the infinite degree of freedom enabled by the soft material, tactile sensors for the rigid gripper might not be suitable for the soft gripper. The choice of available tactile sensors is limited. (2) Unlike the rigid gripper, whose motion and force feature are the main concern, the soft gripper also needs to consider the large deformation feature.

In the research community, there are some prior attempts to develop soft sensors for tactile recognition. She et al. combined a resistive flexible sensor with an SMA driver for curvature detection and feedback.\textsuperscript{29} Chossat et al. applied ionic and liquid metals to develop highly flexible strain sensors.\textsuperscript{30} A flexible “skin” sensor that could identify pressure and strains independently was invented.\textsuperscript{31} Similarly, a flexible and extensible capacitive sensor was designed by Li et al.,\textsuperscript{32} and a soft optical sensor for measuring fingertip contact forces was proposed by Cho et al.\textsuperscript{33} In all these works, soft sensors are designed for specific soft grippers with specialized materials and structures. They are expensive and might not be applicable to most of the soft grippers, thus hard to be utilized in the practical grasping tasks. A more promising and more efficient solution is employing the existing sensors to the recognition of soft grippers.

In the application of existing tactile sensors to the soft grippers, Homberg et al.\textsuperscript{34} were the first ones to use bending sensors for a haptic recognition that provides configuration estimations to distinguish among a set of objects. Gandarias et al.\textsuperscript{35} used the high-precision array type tactile sensor to detect the tactile images of the two-finger flexible gripper in contact with an object. Chen et al.\textsuperscript{36} embedded the bending sensor into the soft pneumatic gripper and established the relationship between the diameter of the grasping ball and the output value of the bending sensor by curve fitting. In all these methods, plenty of experiments are implemented to capture accurate tactile information. They require strenuous effort if a wider range of objects are intended to be accurately recognized.

Having realized the pros and cons of the existing visual and tactile recognition methods, we come up with an idea to combine the two methods for low-cost, efficient, and accurate object recognition. A visual–tactile recognition method is proposed in this article. Visual recognition is firstly applied to a rough classification of the objects, which allows recognizing objects with obvious features like color and shape. Tactile recognition is then applied to achieve accurate identification by further assessing the property of the object, such as size and weight. To elaborate on the proposed recognition method, a self-developed soft gripper is adopted as the study object. The organization of the article is as follows. The second section briefly describes the structure and fabrication of the soft gripper, where the camera and embedded tactile sensors are introduced. The third section summarizes the proposed visual–tactile recognition method, followed which the vision recognition based on a faster RCNN\textsuperscript{37} algorithm, and the tactile recognition based on a machining learning algorithm is illustrated in the fourth and fifth sections, respectively. The sixth section introduces the control system. The experiments are given in the seventh section before the conclusions are drawn in the eighth section.

**Soft gripper and adopted sensors**

Figure 1 shows a grasping robotic system consisting of an articulated serial robot and a soft gripper. The articulated serial robot is a UR3 robot\textsuperscript{38} adopted for changing the position and orientation of the gripper. A Kinect v2\textsuperscript{39} is selected as the visual sensor. As is shown in the figure, the Kinect v2 consists of a color camera, an infrared camera, and an infrared transmitter. The color camera is to obtain the RGB image of the view. The infrared transmitter emits infrared (IR), which will reflect if it touches the surface of the object. The reflected IR will be captured by the infrared camera. Judging by the time when to receive the reflected IR, the depth image of the object is formed.

The proposed soft gripper has three fingers, which are actuated by the pneumatic actuators. As shown in Figure 2, each finger has an actuation part enabling the bending of the finger. The actuation part is composed of a multichamber
and an inextensible layer. The former is made from silicone elastomeric material (Dragon Skin 30, Smooth-On Inc., Macungie, Pennsylvania, USA) and the latter is fiberglass mesh. By pumping air into the chambers, the pressure caused by the inflation of the chambers results in the bending of the fiberglass mesh. Considering the future tactile recognition, a perception part is designed to integrate the tactile sensors when fabricating the finger, in which bending sensors and pressure sensors are embedded.

As shown in Figure 2(b), Spectra Symbol is selected to be the bending sensor, which is applied to capture the bending information of the finger. The Spectra Symbol deforms along with the fiberglass mesh. Its bending angle is converted to the change of resistance. The length of it can reach up to 95.25 mm, which is long enough to measure the bending of the finger. It has good flexibility and thus is suitable to be attached to the surface of soft material. Figure 2(c) shows the chosen pressure sensor FSR 402 by Interlink Electronics Inc., Westlake Village, Southern California that obtains the force of two contacting surfaces during object grasping. The resistance of the force sensor decreases as the contacting force increases. The thin and bendable structure allows it to be embedded in the soft material.

The sensors are calibrated by experiments before embedding in the soft finger. The force sensors are calibrated by the strain dynamometer, where the output voltage of the force sensor is measured, as shown in Figure 3(a). With the increasing of the contact force, the resistance decreases and the output voltage increases accordingly. By exerting forces between 0 N and 10 N onto the pressure sensor, the corresponding voltage is measured. A cubic polynomial function is applied to fit the curve between the force and voltage as

\[
V = 0.0008|F|^3 - 0.025|F|^2 + 0.2707|F| + 0.0088
\]

The input of the bending sensor is air pressure and the output is the voltage. When given certain air pressure, the soft finger bends and the curve is drawn, as shown in Figure 3(b). The voltage corresponding to this air pressure is measured. Hence, the bending information of the soft finger is demonstrated by the relationship between air pressure and voltage. The range for the air pressure is 0–60 kPa and for the output voltage is 1.74–2.26 V. As the air pressure increases, the voltage decreases.

The soft finger is made by casting. Instead of assembling sensors and the soft finger after fabrication, the sensors are
directly embedded into the soft fingers during fabrication. As shown in Figure 4, the molds are 3D printed, with which the soft finger is cast step-by-step, that is, firstly, the actuation part and then the embedded sensor part.

The casting of the actuation part is summarized as follows: (1) Assemble molds A and B. Pour in the uncured silicone (Dragon Skin30, Smooth-On Inc., Macungie, Pennsylvania, USA). (2) Place the mold in a vacuum deforming barrel to remove air bubbles generated from the expansion. Heat the mold at 50–60°C for 40 min in a temperate box. (3) Cool down the mold at room temperature. The heating up and cooling down process is to speed up the solidification of silicone.
The casting of the embedded sensor part is shown in the following. (4) Pour the uncured silicone into mold C until it reaches the height 1.2–1.8 mm. This is to cast the base for sensors like the step 1. Repeat the heating up and cooling down process like step 2. (5) Attach the pressure sensors onto the base. Pour in the uncured silicone until the pressure sensor is covered. Repeat the heating up and cooling down process. Attach bending sensors onto the force sensor layer and repeat a similar process as above. On top of the bending sensor layer, put the fiberglass mesh and repeat the process as above. (6) The actuation part and the embedded sensor part are finally connected by the casting of uncured silicone.

**Visual–tactile object recognition method**

As mentioned above, visual and tactile recognition methods have their own pros and cons. Visual recognition is fast in localization, but accurate objective identification requires a camera with high performance and some features are difficult to be captured. Tactile recognition is precise in collecting the features and identifying the object. However, soft fingers are required to have one or more contacts with an unknown target object to obtain tactile information. The above entire process reduces object recognition efficiency. Inspired by the human perception process, which identifies object firstly by vision and then by touching and grasping, we propose a visual–tactile fusion recognition method for efficient and practical object recognition.

As shown in Figure 5, visual recognition is firstly applied for object localization and initial classification, which is realized by the depth image and the RGB image obtained by Kinect v2, respectively. A region-based algorithm called faster RCNN is selected to extract the object features from the images and then classify these features by a classifier. If the object cannot be recognized from the RGB image, the characteristics obtained from vision recognition are not enough to identify the object. Due to the lack of information, similar objects would fail to be recognized. We define such objects as attribute missing categories, for instance, balls with the same color but different sizes or the
same bottles with different volumes of water. To solve this problem, the tactile recognition is then applied.

For the tactile recognition, \( N \) sets (rough within the range 70–100) of object grasping experiments are implemented to collect tactile information from the embedded force and bending sensors. The tactile information is stored in a vector of tactile features. The “bagged trees” algorithm is adopted as a classifier. Randomly, select \( n_1 \) sets of experimental data to perform the data training and cross verification. The \( n_2 = N - n_1 \) sets of experimental data are applied to validate the classification models. With the aid of trained models, more features of the objects are captured. Usually, one classifier corresponds to one tactile feature. For instance, the classifier 1 relates to the size of the object, and the classifier 2 is attributed to the weight. By assigning one feature to one classifier, each classifier requires less training data to classify objects. Thus, fewer experiments are needed and the efficiency of information collection improves. It is worth mentioning that the tactile features closely relate to the accuracy of the object recognition. The more quantity and better performance of the tactile sensors, the higher accuracy of the recognition results. In our work, force and bending sensors are adopted to compromise between accuracy and practical fabrication. More tactile sensors can be embedded to the soft fingers to collect more information if another soft gripper is employed and more complicated application scenarios are considered.

Herein, visual and tactile recognitions deal with their own data separately by a decision-making step. The results from the visual and tactile recognition are fused on the decision-making layer as the final recognition conclusion. The object is expected to be accurately recognized by the combination of the two recognition methods.

**Visual recognition based on faster RCNN**

As has mentioned, Kinect v2 is applied to obtain the depth and RBG images of the object. Since the information on the location of the object is also aimed to be acquired for the grasping, a target algorithm called faster RCNN\(^37\) is selected to address the problem of identification, classification, and localization.

In the faster RCNN model, the image is put into the convolutional neural network to generate a feature map. A region proposal network is then applied to come up with feature squares, from which the detail features are identified by a region of interest. The features are classified and the visual recognition results can be obtained. The details are carried out by the object detection API from Google as follows.

1. Derive the images of the object. Label it by the *Labelling* and transfer it into TensorFlow format.
2. Load the data training model *Faster RCNN Inception ResNet v2*. Modify the training model and create the explanation file.
3. Train the data by *TensorFlow* and output the model. The GPU employed during training is the *NVIDIA GeForce GTX 1080*.
4. Input the test image and obtain the recognition results. Compare with the real object and assess the accuracy.
5. Repeat steps (1)–(4) and compute the average recognition accuracy.

To show the effectiveness of faster RCNN, another well-known visual recognition algorithm called SSD\(^43\) is applied. The visual recognition based on SSD is similar to the procedure above. The difference is in the second step, where the model for data training in the SSD algorithm is SSD MobileNet v1.

**Tactile recognition based on machining learning**

The tactile recognition process can be divided into two steps. First of all, object grasping experiments are implemented. Data from the bending and force sensors are collected, from which the tactile features are extracted. Then, these tactile features are classified by the machining learning algorithms and the grasping object can be recognized.

**Data collection**

The process of tactile data collection is provided in Table 1. An object is handed to the soft gripper by the operator with different orientations and positions for better recognition robustness. The soft gripper gradually touches and grasps the object. After the object is successfully lifted, the soft gripper remains still for \( \sim 5–10 \) s. The data from the sensors are recorded and stored in a matrix \( \Theta_i \), where \( i \) represents \( i \)th grasp experiments. The figure of the data in matrix \( \Theta_i \) is drawn, from which the steady state is recognized, and the corresponding values are kept in a vector \( o_i \).

An example is given to illustrate the data collection process as follows. Ten pressure sensors and two bending sensors were embedded in the soft fingers. After holding the
object 5–10 s, the data acquisition card started collecting the sensors’ data. The acquisition frequency is \(20–50\) Hz. The data were stored in the matrix \(Y_{1}\), whose dimension is \(s/C_2\), shown in Figure 6, where \(s\) is the number of frames in the haptic sequence. The middle 100–150 values were roughly constant. This period is selected as a steady state, as shown between the dotted lines. The same experiment was repeated 70–100 times for different grasping poses, and the average of the values at steady state was calculated. Finally, the tactile feature vector \(o_1\) is obtained as
\[
\begin{align*}
o_1 = [0.0007, 1.2634, 0.6488, 0.0005, 0.0009, 1.5498, & \\
0.0004, 0.0014, 0.0003, 0.0001, 0.0008, 2.169]
\end{align*}
\]

**Machine learning algorithms**

After obtaining the tactile feature of the objects, machining learning algorithms are adopted to train and classify the objects. We applied five different machining learning algorithms, as given in Table 2. The decision tree algorithm \(^4^4\) is easy to be implemented, in which three decision tree classifiers are used in feature processing. For the discriminant analysis, linear and quadratic discriminant analysis classifiers are studied. Support vector machine (SVM) \(^4^5\) has been widely used in various classification problems and has achieved good results. Herein, six different SVMs are used in feature processing. The K-nearest neighbor (KNN) algorithm \(^4^6\) has also been widely used in current tactile recognition, and six KNN classifiers are employed for comparative analysis. In addition, some classifiers integrated from different algorithms have good behavior in classification problems. \(^4^7\) Five ensemble classifiers \(^4^8\) are adopted. In total, 22 different classifiers are applied for analysis, from which the one with the best accuracy will be selected for fusion recognition.

**Experimental verification**

**Visual recognition**

In the visual recognition experiment, 20 different objects \(^2^8\) that are commonly seen in daily life are selected, as shown in Figure 7 and Table 3. The objects are in different shapes, colors, sizes, and weights. Especially, the shape of the balls (C14, C15, C16, and C17) is the same but they are with different sizes (diameters are 63, 83, 98, and 120 mm, respectively).
respectively). The color of C16 and C17 is the same but they are in different colors compared with C14 and C15. In addition, to test the recognition accuracy of the objects with the same shapes and color but different weights, the same bottle with different volumes of water is set. They are C18, C19, and C20.

In the experiment, 2700 images of the objects are taken, in which 1800 images are used for training models and the remaining 900 images for verifications. The recognition results are provided in Table 4. Both models show acceptable accuracy if the objects are with distinguished shape and color (C1–C13, for instance). For the balls, C14 and C15 are well recognized but C16 and C17 could not be recognized by both models. For the bottles, the recognition accuracy of the SSD model is slightly higher than the faster RCNN model. However, neither of them reaches acceptable level. In conclusion, visual recognition is weak for the objects in different sizes and weights. Between these two models, the average recognition accuracy of the faster RCNN model is 80.78% (727/900), which is higher than the SSD model. Therefore, faster RCNN is chosen as the visual algorithm, and the tactile information is necessary for better recognition.

**Tactile recognition**

Tactile experiments are carried out on the same objects, as given in Table 3 and Figure 7. Each object is grasped by the soft gripper 70–100 times (see Figure 8), among which data of 50–70 times of grasping are regarded as the training set and the rest is the testing set. The training set is for the data training by machining learning algorithms given in Table 2.

![Three-finger gripper: (a) open state and (b) grasping state.](image)

To avoid overfitting, a 10-fold cross-validation method is applied to access the algorithms. The recognition accuracy is given in Table 5. Among the 22 classifiers of different machining learning algorithms, the average recognition accuracy of the bagged trees is the highest (87.8%).

The testing set is applied to further access the recognition of the bagged trees. An object category label can be obtained during the data training by the classifier on the training set. The value of this label is named as the predicted value. Similarly, an actual category label is defined by the testing set, of which the value is named as the actual value. These two labels are adopted to the 10-fold cross-validation method and a confusion matrix is generated. As shown in Figure 9, the horizontal axis denotes the predicted value and the vertical axis represents the actual value. The diagonal elements show the probability that the predicted

| Code | Item name | Code | Item name | Code | Item name | Code | Item name |
|------|-----------|------|-----------|------|-----------|------|-----------|
| C1   | Apple     | C6   | Mug       | C11  | Paper cup | C16  | Ball 3 (d = 98 mm) |
| C2   | Orange    | C7   | Paint box | C12  | Stapler   | C17  | Ball 4 (d = 120 mm) |
| C3   | Square milk box | C8   | Glasses case | C13  | Wire     | C18  | Empty bottle (14.7 g) |
| C4   | Radio     | C9   | Conditioner | C14  | Ball 1 (d = 63 mm) | C19  | Half full bottle (259.3 g) |
| C5   | Pencil sharpener | C10  | Tea cup   | C15  | Ball 2 (d = 83 mm) | C20  | Full bottle (422.6 g) |

| Objects (C1–C10) | Accuracy of SSD (%) | Accuracy of faster RCNN (%) | Objects (C11–C20) | Accuracy of SSD (%) | Accuracy of faster RCNN (%) |
|-------------------|----------------------|-----------------------------|-------------------|----------------------|-----------------------------|
| C1                | 91.1                 | 100                         | C11              | 95.6                 | 100                         |
| C2                | 97.8                 | 100                         | C12              | 84.4                 | 100                         |
| C3                | 100                  | 100                         | C13              | 100                  | 100                         |
| C4                | 88.9                 | 100                         | C14              | 93.3                 | 100                         |
| C5                | 86.7                 | 97.8                        | C15              | 100                  | 100                         |
| C6                | 86.7                 | 95.5                        | C16              | 46.7                 | 13.3                        |
| C7                | 100                  | 100                         | C17              | 33.3                 | 26.7                        |
| C8                | 91.1                 | 100                         | C18              | 73.0                 | 17.8                        |
| C9                | 100                  | 97.8                        | C19              | 46.7                 | 6.7                         |
| C10               | 91.1                 | 100                         | C20              | 53.3                 | 60.0                        |
and actual values are the same. By analyzing the confusion matrix, it is summarized that the average recognition accuracy of the bagged tree reaches 88.76% (545/614), among which apple (C1) and orange (C2), pencil sharpener (C5), and conditioner (C9) are 100%. In the visual recognition experiment, the objects with the same shape but different weights (C18, C19, and C20) failed to be recognized. In the tactile recognition experiment, the recognition accuracies of C19 and C20 are up to 91% and 94%, indicating that tactile recognition can distinguish objects with different weights. However, the recognition accuracy of C18, as well as the cup (C6) and ball 1 (C14), is less than 70%. The reason for this result might be that the information from the tactile sensors is not rich enough to recognize the objects with similar features, especially the objects with similar shapes, sizes, and weights.

Visual–tactile recognition

As has shown by the visual and tactile recognition experiments in “Visual recognition” section and “Tactile recognition” section, the visual recognition can efficiently distinguish objects with varied shapes and colors, but it is weak for the objects in different sizes and weights. The tactile recognition can solve the problem of size and weight. However, a lot of grasping experiments are necessary to collect enough information for better accuracy. To take full advantage of both the visual and tactile recognition methods, the visual–tactile recognition method is proposed in this article and the experiment is carried out in this section.

The same 20 objects are adopted again by the visual–tactile recognition experiment. As shown in Figure 9, tactile recognition is implemented to compensate for the missing attributes of objects after the visual recognition.

Table 5. Accuracy of classifiers on the training data.

| Classifiers (1–11) | Accuracy (%) | Classifiers (12–22) | Accuracy (%) |
|--------------------|--------------|---------------------|--------------|
| Complex tree       | 76.9         | Fine KNN            | 86.2         |
| Medium tree        | 50.4         | Medium KNN          | 79.6         |
| Simple tree        | 21.6         | Coarse KNN          | 38.6         |
| Linear discriminant| 62.0         | Cubic KNN           | 78.3         |
| Quadratic discriminant | 79.3    | Cosine KNN          | 77.6         |
| Linear SVM         | 81.9         | Weighted KNN        | 84.5         |
| Quadratic SVM      | 87.1         | Boosted trees       | 73.6         |
| Cubic SVM          | 86.0         | Bagged trees        | 87.8         |
| Fine Gaussian SVM  | 62.6         | Subspace            | 56.6         |
|                    |              | discriminant        |              |
| Medium Gaussian SVM| 84.4         | Subspace KNN        | 84.4         |
| Coarse Gaussian SVM| 62.1         | RUSBoosted Trees    | 66.5         |

KNN: K-nearest neighbor; SVM: support vector machine. The bold values indicate the classifier has the highest accuracy.

Figure 9. Confusion matrix of the bagged tree.
Especially, the ball 3 (C16) and the ball 4 (C17) with the same color differ only in size. The empty bottle (C18), the half-full bottle (C19), and the filled bottle (C20) differ only in weight. The two groups of objects were difficult to recognize by visual and needed to be recognized by tactile. The trained faster RCNN model was used for visual recognition. Two tactile classifiers were used to recognize different sizes and qualities.

As shown in Figure 10, visual recognition is firstly implemented. If the object can be identified, the

Figure 10. Experiments of the visual–tactile recognition (tactile classifier 1:3: identify objects of different sizes; tactile classifier 2:4: identify objects with different weights).

Figure 11. Confusion matrix of three-fingers soft gripper visual–tactile fusion.
Table 6. Accuracy comparisons of different recognition methods.

| Objects (no.)            | Visual | Tactile | Visual–tactile |
|--------------------------|--------|---------|---------------|
| Balls with different sizes (2) | 20%    | 86.36%  | 97.73%        |
| Bottles with different weight (3) | 28.15% | 86.29%  | 95.97%        |
| All objects (20)         | 80.78% | 88.76%  | 98.70%        |

The accuracies of recognizing the 20 objects are accessed by the confusion matrix shown in Figure 11. On average, the accuracy is 98.70%, showing a good recognition result. Ball 3 (C16) and ball 4 (C17) fail to be recognized by the visual method. By combining the information from vision and tactile classifiers 1 and 3, the recognition accuracy of C16 and C17 is 100%. Similarly, the empty bottle (C18), the half-full bottle (C19), and the filled bottle (C20) cannot be identified by the visual recognition method. After recognized by the visual–tactile fusion method, the recognition accuracies of C18, C19, and C20 are all 100%. The recognition accuracy of balls and bottles is greatly improved compared with the accuracy of the tactile recognition method only. This might contribute to the combination of the information from both visual and tactile sensors. It shows that the visual–tactile fusion method at the decision-making level can make full use of visual and tactile recognition methods and improved the recognition accuracy of objects. The comparisons on the accuracy of different recognition methods are provided in Table 6. The average accuracy of visual, tactile, and visual–tactile recognition methods is 80.78%, 88.76%, and 98.7%, respectively, indicating that the best recognition results can be obtained by the visual–tactile recognition. For the objects with the same shape but different sizes, the accuracy of visual recognition is only 20%. It improves a lot by the tactile recognition, whose accuracy reaches 86.36%. However, the best accuracy is achieved by the visual–tactile recognition that increases 11.37% compared with the tactile recognition. Similarly, for the objects with the same shape and color but different weights, the recognition accuracy of visual recognition is the lowest (28.15%), followed by the tactile recognition (86.29%), and the highest accuracy is from visual–tactile recognition (95.97%). The results show that the visual–tactile recognition method can identify daily objects with high accuracy.

Conclusions

A visual–tactile recognition method is proposed to efficiently and accurately identify the unknown object for a successful grasping of the soft gripper. A three-step procedure is presented, including initial recognition by vision based on the faster RCNN model, detail recognition by touch based on machining learning algorithm, and data fusion at decision-making level.

Considering the visual and tactile sensors, the design and fabrication of the soft gripper are first implemented. A Kinect v2 is adopted, and the RGB and depth images of the object can be collected. Bending sensors and pressure sensors are calibrated and embedded into the soft finger during fabrication, which turns the bending and contacting forces into resistance. For the initial recognition by vision, the faster RCNN is applied for classification and localization of the object. The identified results will be directly regarded as the final result if the object does not involve size or weight and can be fully recognized. If not, the detail recognition by touch is carried out. Machining learning algorithms are adopted to train grasping data. The information from both vision and tactile is finally combined at the decision-making layer, and the output is the recognition result. Experiments are implemented to verify the proposed method. The average accuracy of the proposed method is higher than visual recognition and tactile recognition, confirming the feasibility of the visual–tactile recognition.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research work was supported by the National Natural Science Foundation of China (NSFC) under grant no. 51675366 and Tianjin Technology and Science Plan Project under grant nos 18YFSDZC00010 and 18YFZCSF00590.

ORCID iD

Binbin Lian https://orcid.org/0000-0002-6747-8565

References

1. Wang YZ, Gupta U, Parulekar N, et al. A soft gripper of fast speed and low energy consumption. Sci China Technol Sci 2019; 62(1): 31–38.
2. Hughes J, Culha U, Giardina F, et al. Soft manipulators and grippers: a review. Front Robot AI 2016; 3: 69.
3. Rus D and Michael T. Design, fabrication and control of soft robots. Nature 2015; 7553: 467–475.
4. Laschi C, Mazzolai B, and Cianchetti M. Soft robotics: technologies and systems pushing the boundaries robot abilities. Sci Robot 2016; 1(1): eaah3690.
5. Udupa G, Sreedharan P, Dinesh PS, et al. Asymmetric bellow flexible pneumatic actuator for miniature robotic soft gripper. J Robot 2014; 2014: 11.

6. Low JH, Cheng N, Khin PM, et al. A bidirectional soft pneumatic fabric-based actuator for grasping applications. In: 2017 IEEE/RSJ international conference on intelligent robots and systems (IROS). Vancouver, BC, Canada, 24–28 September 2017, pp. 1180–1186. IEEE.

7. Sun T, Chen Y, Han T, et al. A soft gripper with variable stiffness inspired by pangolin scales, toothed pneumatic actuator and autonomous controller. Robot Comput Integr Manuf 2020; 61: 101848.

8. Rodrigue H, Wang W, Kim D-R, et al. Curved shape memory alloy-based soft actuators and application to soft gripper. Compos Struct 2017; 176: 398–406.

9. Wang W and Ahn S-H. Shape memory alloy-based soft gripper with variable stiffness for compliant and effective grasping. Soft Robot 2017; 4(4): 379–389.

10. Yang Y, Chen Y, Li Y, et al. Bioinspired robotic fingers based on pneumatic actuator and 3D printing of smart material. Soft Robot 2017; 4(2): 147–162.

11. Yang Y, Chen Y, Li Y, et al. Novel variable-stiffness robotic fingers with built-in position feedback. Soft Robot 2017; 4(4): 338–352.

12. Rosset S, Araromi OA, Shintake J, et al. Model and design of dielectric elastomer minimum energy structures. Smart Mater Struct 2014; 23(8): 085021.

13. Shian S, Bertoldi K, and Clarke DR. Dielectric elastomer based “grippers” for soft robotics. Adv Mater 2015; 27(43): 6814–6819.

14. Shintake J, Rosset S, Schubert B, et al. DEA for soft robotics: 1-gram actuator picks up a 60-gram egg. Proc SPIE 2015; 9430: 94301S.

15. Shintake J, Rosset S, Schubert B, et al. Versatile soft grippers with intrinsic electroadhesions based on multifunctional polymer actuators. Adv Mater 2016; 28(2): 231–238.

16. Yap HK, Ng HY, and Yeow C-H. High-force soft printable pneumatics for soft robotic applications. Soft Robot 2016; 3(3): 144–158.

17. Zhu M, Mori Y, and Wakayama T. A fully multi-material three-dimensional printed soft gripper with variable stiffness for robust grasping. Soft Robot 2019; 6(4): 507–519.

18. Polygerinos P, Wang Z, Kevin C, et al. Soft robotic glove for combined assistance and at-home rehabilitation. Robot Auto Syst 2015; 73: 135–143.

19. In H, Kang BB, Sin M, et al. Exo-glove: a wearable robot for the hand with a soft tendon routing system. IEEE Robot Auto Mag 2015; 22(1): 97–105.

20. Vogt DM, Park Y-L, and Wood RJ. Design and characterization of a soft multi-axis force sensor using embedded microfluidic channels. IEEE Sens J 2013; 13(10): 4056–4064.

21. Yu Y, Gu J, Kl Mann G, et al. Development and evaluation of object-based visual attention for automatic perception of robots. IEEE Trans Auto Sci Eng 2012; 10(2): 365–379.

22. Liu H, Wu Y, Sun F, et al. Recent progress on tactile object recognition. Int J Adv Robot Syst 2017; 14(4): 1–12.

23. Kaboli M, Walker R, and Cheng G. Re-using prior tactile experience by robotic hands to discriminate in-hand objects via texture properties. In: 2016 IEEE international conference on robotics and automation (ICRA). Stockholm, Sweden, 16–21 May 2016, pp. 2242–2247. IEEE.

24. Kerr E, McGinnity TM, and Coleman S. Material recognition using tactile sensing. Exp Syst Appl 2018; 94: 94–111.

25. Yi Z, Calandra R, Veiga F, et al. Active tactile object exploration with gaussian processes. In: 2016 IEEE/RSJ international conference on intelligent robots and systems (IROS). Daejeon, South Korea, 9–14 October 2016, pp. 4925–4930. IEEE.

26. Regoli M, Jamali N, Metta G, et al. Controlled tactile exploration and haptic object recognition. In: 2017 18th international conference on advanced robotics (ICAR). Hong Kong, China, 10–12 July 2017, pp. 47–54. IEEE.

27. Anghinolfi D, Cannata G, Mastrogiavanni F, et al. On the problem of the automated design of large-scale robot skin. IEEE Trans Auto Sci Eng 2013; 10(4): 1087–1100.

28. Liu H, Yu Y, Sun F, et al. Visual–tactile fusion for object recognition. IEEE Trans Auto Sci Eng 2016; 14(2): 996–1008.

29. She Y, Li C, Cleary J, et al. Design and fabrication of a soft robotic hand with embedded actuators and sensors. J Mech Robot 2015; 7(2): 021007.

30. Chossat JB, Park Y-L, Wood RJ, et al. A soft strain sensor based on ionic and metal liquids. IEEE Sens J 2013; 139: 3405–3414.

31. Park YL, Chen B-R, and Wood RJ. Design and fabrication of soft artificial skin using embedded microchannels and liquid conductors. IEEE Sens J 2012; 12(8): 2711–2718.

32. Li S, Zhao H, and Shepherd RF. Flexible and stretchable sensors for fluidic elastomer actuated soft robots. MRS Bullet 2017; 42(2): 138–142.

33. Cho H, Lee H, Kim Y, et al. Design of an optical soft sensor for measuring fingertip force and contact recognition. Int J Control Auto Syst 2017; 15(1): 16–24.

34. Homberg BS, Katzschmann RK, Dogar MR, et al. Haptic identification of objects using a modular soft robotic gripper. In: 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS). Hamburg, Germany, 28 September 2015–2 October 2015, pp. 1698–1705. IEEE.

35. Gandarias J, Gómez-de-Gabriel J, and García-Cerezo A. Enhancing perception with tactile object recognition in adaptive grippers for human–robot interaction. Sensors 2018; 18(3): 692.

36. Chen Y, Guo S, Li C, et al. Size recognition and adaptive grasping using an integration of actuating and sensing soft pneumatic gripper. Robot Auto Syst 2018; 104: 14–24.

37. Ren S, He K, Girshick R, et al. Faster r-cnn: Towards realtime object detection with region proposal networks. Proc 28th Int Conf Neural Inf Process 2015: 91–99.

38. http://www.allied-automation.com/partners/universal-robots/ur3-robot/.

39. https://blog.csdn.net/hehedadaq/article/details/80557926.

40. https://www.spectrasymbol.com/.
41. https://www.trossenrobotics.com/productdocs/2010-10-26-
DataSheet-FSR402-Layout2.pdf.
42. Lahat D, Adali T, and Jutten C. Multimodal data fusion: an
overview of methods, challenges, and prospects. Proc IEEE
2015; 103(9): 1449–1477.
43. Liu W, Anguelov D, Erhan D, et al. SSD: single shot multi-
box detector. In: European conference on computer vision,
Amsterdam, The Netherlands, 11–14 October 2016, pp.
21–37. Berlin: Springer.
44. Chitta S, Piccoli M, and Sturm J. Tactile object class and
internal state recognition for mobile manipulation. In: 2010
IEEE international conference on robotics and automation.
Anchorage, AK, USA, 3–7 May 2010, pp. 2342–2348. IEEE.
45. Wu X, Sun C, Zou T, et al. SVM-based image partitioning for
vision recognition of AGV guide paths under complex
illumination conditions. Robot Comput Integr Manuf 2020;
61: 101856.
46. Drimus A, Kootstra G, Bilberg A, et al. Design of a flexible
tactile sensor for classification of rigid and deformable
objects. Robot Auto Syst 2014; 62(1): 3–15.
47. Kohavi R. A study of cross-validation and bootstrap for accu-
rate estimation and model selection. In: 14th International
Joint Conference on Artificial Intelligence, Morgan Kauf-
mann 1995; 14(2): 1137–1145.
48. Rahman A and Tasnim S. Ensemble classifiers and their
applications: a review. Int J Comput Trends Technol 2014;
10(1): 31–35.
49. Zhang W, Sun F, Wu H, et al. A framework for the fusion of
visual and tactile modalities for improving robot perception.
Sci China Inf Sci 2017; 60(1): 012201.