

A Human Manipulator Collaboration-based Scheme for Object Inspection by Robust and Fast Recognition of Hand Pose Sensing Images of Numeric Symbols Combined with Non-numeric Expressions for Smart Factories

Ing-Jr Ding, * Meng-Chuan Hsieh, and Zhi-Xuan Wu

Department of Electrical Engineering, National Formosa University,
No. 64, Wunhua Rd., Huwei Township, Yunlin County 632, Taiwan, ROC

(Received August 25, 2022; accepted February 24, 2023)

Keywords: automatic optical inspection, human–robot collaboration, manipulator, image sensor, hand pose recognition, smart factory

The Fourth Industrial Revolution, or Industry 4.0, has among its goals improving efficiency and increasing throughput. To achieve these goals, the smart factory for Industry 4.0 requires digital techniques of artificial intelligence (AI), image signal processing (ISP), and advanced robotics that can support the system integration of software and hardware. An interactive mode of human–robot collaboration (HRC) is a significant focus of the smart factory to increase the intelligence of robots and further promote autonomous systems. In this study, we have developed a scheme for object inspection based on HRC. Different from automatic optical inspection (AOI) without the involvement of an operator, the inspection scheme developed significantly increases interactions between the operator and the manipulator. In the proposed HRC-based object inspection, the manipulator is viewed as an assistant to the operator to deliver the inspected object by grabbing, transporting, and releasing it. When the manipulator moves the object to the side of the operator, the operator then makes a detailed inspection on the basis of his expert knowledge; after completing the inspection, the operator executes consecutive hand poses to “tell” an image sensor located in the manipulator the results of the inspection, including both the object quality rank and the object material type. Hand pose recognition is used to classify the hand poses of the operator. In this study, a two-phase hand pose recognition strategy was adopted on the basis of simple finger segmentations. A robust and fast recognition approach that is capable of making classifications of hand poses of numeric symbols combined with non-numeric expressions was developed. The HRC-based object inspection was carried out in a laboratory environment that simulated a real system. Experiments on HRC-based object inspection with classifications of five hand poses of non-numeric expressions and nine hand poses of numeric symbols demonstrated both the effectiveness and efficiency of this approach.

*Corresponding author: e-mail: eugen.ding@gmail.com
<https://doi.org/10.18494/SAM4236>

1. Introduction

With the proper incorporation of artificial intelligence (AI) techniques, conventional factories (or manufacturing sites) can be significantly improved. The smart factory with additional AI can make a product line more efficient and greatly increase throughput.⁽¹⁾ Typical factory automation that employs various devices, such as manipulators (also known as mechanic arms), vehicles (carrier platforms that move), and conveyor belts, is frequently in current use, but the inclusion of AI in these systems is rarely considered. To promote automation in factories, various types of sensor combined with different AI process techniques are being rapidly imported into the field to upgrade typical automatic devices, e.g., a manipulator can be enhanced with machine vision by the addition of a CMOS image sensor (CIS) or a charge coupled device (CCD)-based image sensor,⁽²⁾ an automatic guided vehicle (AGV) can be equipped with ultrasonic sensors,⁽³⁾ or an autonomous mobile robot (AMR) can be fitted with advanced sensors for light detection and ranging (LiDAR), multiple image sensors, or other types of sensor.⁽⁴⁾

For a manipulator with machine vision, automatic optical inspection (AOI) is a practical and specific application.⁽⁵⁾ The typical operation in AOI is that the object quality is evaluated by processing the object's image from a sensor generally deployed above the object to provide a top-to-down view. In typical AOI operations, obtaining an image with high resolution and making a detailed analysis of that image require expensive image sensors. A series of studies discussing this topic have appeared in recent years,^(6–10) and the image processing and recognition of the target object to be inspected have been major concerns. Reshadat and Kapteijns⁽⁶⁾ reported that, to alleviate the problem of high false rates occurring in cases in which a component was incorrectly identified as a 'fail,' various classical machine learning-based classifiers, such as a decision tree (DT), random forest (RF), K-nearest neighbors (KNNs), and artificial neural network (ANN), can be trained and used to visually inspect a printed circuit board (PCM). In other studies, machine-learning-based AOIs were adapted to the deep learning strategies; the deep learning model of convolutional neural networks (CNNs) was used for defect inspection tasks where fully connected layers at the terminus of the CNN structure were replaced by layers of convolutional computations for possible performance improvements in the classifications;⁽⁷⁾ deep-learning-based detection approaches were applied to establish an AOI system for performing defect detection in ceramic substrates for which three different models with integrated structures, "ResNeXt+YOLO v3," "Inception v3+YOLO v3," and "YOLO v3," were evaluated.⁽⁸⁾ The technique of "3D AOI with At-Line X-ray" and related 3D AOI promotions were presented as fast and advanced solutions to overcome the problem of "false fails and escapes" in an in-line inspection environment.⁽⁹⁾ An AOI system used to perform inspections and classifications of surface defects in the glass of touch panels by the feature extraction of Hu's moment invariants and classifying computations of back-propagation-based ANN models has also been developed.⁽¹⁰⁾

In all these AOI-related studies,^(6–10) the target for imaging by the sensor, i.e., the region of interest (ROI) in the acquired image, was the object being inspected. From the viewpoint of smart manufacturing with human-robot collaboration (HRC), such AOI systems clearly lack interactions between the operator and the machine and also introduce the serious problem of long idle times for the operator. Different from previously proposed AOI systems, to reach high

efficiency for HRCs, we developed in this work an advanced object inspection system for a smart factory incorporating an HRC-based scheme involving the recognition of operator hand poses. The recognition of operator hand poses in an HRC-based inspection system is fast and robust and can include the classification of hand pose sensing images of numeric symbols combined with non-numeric expressions. The utilization of hand gesture recognition to analyze and then classify a person's hand gestures or hand poses has been achieved in many fields,^(11–15) such as the robotic handling of smart materials,⁽¹¹⁾ the recognition of the intention of hand gestures,⁽¹²⁾ and the recognition of the identity and emotion associated with hand gestures.^(13,14) Studies on the use of hand gesture recognition in object inspections are extremely rare. The main contributions of the proposed HRC-based inspection system with hand pose recognition described in this study are as follows.

- (1) Different from the current AOI of a manipulator with machine vision using a fixed deployed camera, a highly efficient scheme for object inspection based on human–manipulator collaboration, called HRC-based object inspection, is presented for use in a smart factory.
- (2) Compared with the conventional AOI that adopts a fixed deployed camera to capture image data requiring further inspection, the HRC inspection with fast and robust hand pose recognition can provide much more flexibility and competitiveness.
- (3) Compared with the conventional approach to hand pose recognition by finger segmentation proposed in Ref. 15, recognition by finger segmentation in this work is more robust and more adaptable to practical factory or manufacturing scenarios in terms of its ability to tolerate noisy environments with regions of skin-like color that are not part of the hand regions and to recognize “two-phase” hand poses to accomplish the tasks of identification of the object type and categorization of the object quality.
- (4) Compared with the work presented in Ref. 15, the proposed finger segmentation for hand pose recognition can effectively separate images of hand poses conveying numeric symbols (denoting object types) from those of non-numeric expressions (denoting object quality ranks).

2. Object Inspections by HRC with Operator Hand Pose Recognition for Smart Factories

As mentioned, the next-generation smart factory will require considerably more human–machine interaction functionalities designed in online operating machines. Figure 1 illustrates the current AOI method to achieve automatic object inspection by using only an optical camera without the necessity for an operator with specific expertise. In the current AOI scheme, the camera for the image sensing of specific interesting objects is deployed in a fixed location, generally a high position, to be able to perform “top-to-down” acquisitions of the overall scene image, including interesting object regions to be inspected. However, in such a typical AOI scheme as that depicted in Fig. 1, some technique-related issues need to be resolved:

- (1) Object detection algorithms with high performance required: before performing object quality evaluations, all of the specific object regions of interest in the captured scene from the deployed camera need to be detected and segmented by an additionally developed object detection algorithm.

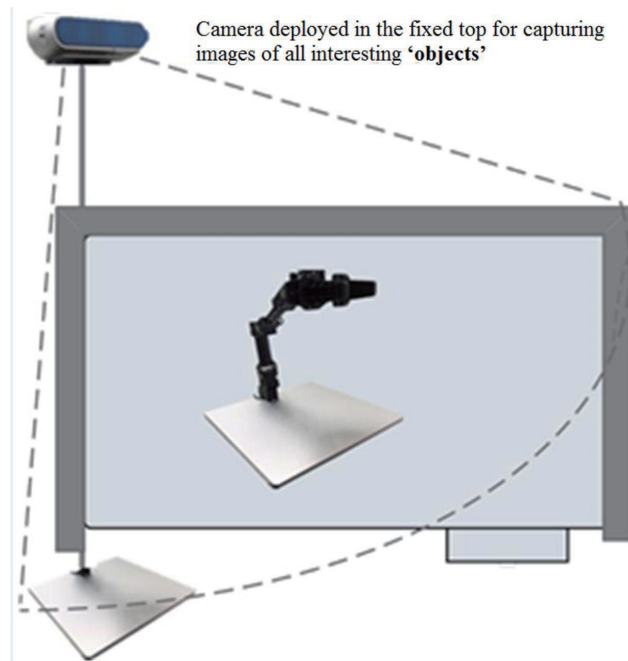


Fig. 1. (Color online) Conventional AOI with machine vision that adopts a camera deployed in the fixed top for up-to-down capturing of all interesting object images for the object quality evaluations (without any participation of the operator).

- (2) Object quality evaluation algorithms with expert-like domain knowledge required: after completing the object detection mentioned above, the ROI of the detected object will be further used for conducting the automatic evaluations on the quality ranks. Such a quality evaluation on the segmented ROI of the detected object will require the development of an additional object quality evaluation algorithm with expert-like domain knowledge.
- (3) Expensive camera with an extremely high-resolution image sensor for the captured images required: for successfully completing object detection and object quality evaluations using the additionally developed approaches mentioned above, the image sensor embedded in the deployed camera will need to be equipped with high specifications to acquire images with an extremely high resolution. In this situation, the image sensor will have to be a type of CCD with high performance, which will increase the hardware cost considerably.

The typical AOI scheme with the manipulators of machine vision alone will inevitably encounter the disadvantage of relatively high costs in the hardware deployments (mainly the high-level image sensor) and software developments (mainly object detection and object inspection algorithms). Different from the conventional AOI object inspection, Fig. 2 illustrates the HRC-based objection inspection scheme developed in this study. As shown in Fig. 2, the object inspection mission will be completed by fine co-operations between the operator and the manipulator. In the presented scheme (see Fig. 2), the camera is appropriately deployed in the manipulator, which will not always be kept in the same position and be accompanied with the moving manipulator to variably make the corresponding movement. It has been noted that in the presented HRC-based object inspection scheme, ROI in the captured image scene will not be the ROI of the object. In contrast, hand poses performed by the operator will be the main interesting

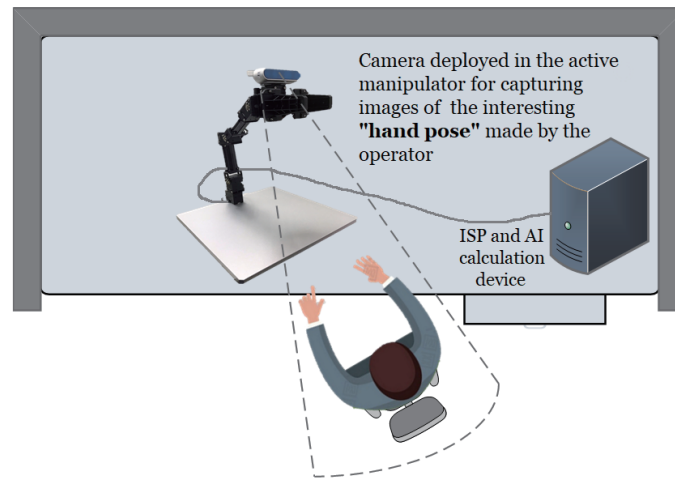


Fig. 2. (Color online) Presented human–machine collaboration-based object inspection scheme where the operator and the manipulator with machine vision (the assistant) cooperate with each other to complete object inspection efficiently (with the operator).

data in sensing image acquisitions of the camera in the manipulator. As shown in Fig. 2, if the manipulator that carries the inspected object moves to the front of the operator, the camera in the manipulator will be ready for capturing the hand pose images acquired by the operator. The hand pose data acquired by the operator with expertise essentially denotes the final inspection result of this object, including the object quality rank and object material type (inspected mainly visually by the operator). The two-phase hand pose recognition calculation will then be performed to obtain the object quality rank (the hand pose recognition result in the first phase) and object material type (the hand pose recognition result in the second phase), which will be explained in detail in the following section. Figure 3 shows the flowchart of the presented HRC-based object inspections, in which, because of the co-operations of the manipulator and operator, the object inspection task will not require expensive computations of object detection and object inspection, and will omit the burden of the expensive image sensor with high specifications. Figure 3 shows that the presented HRC-based object inspection consists mainly of the following five operation steps (containing operations of the manipulator, operator, and computing device):

Step 1 (object grabbing, *manipulator* operations): This step is the starting procedure to make the manipulator correctly grab the specific object.

Step 2 (object moving from the original position to the operator position, *manipulator* operations): When the manipulator completes the procedure of object grabbing, this step will be triggered to ensure that the manipulator moves the inspection object to the operator side for inspection.

Step 3 (object inspection by the operator, *operator* operations): If the inspection object is moved to the operator side (in front of the operator), the operator with expertise will start to check both the quality rank and material type of this inspection object. In this step, the operator will mainly use his eyesight (or sometimes hand touching, nose smelling, and tool measuring) and his expert domain knowledge to complete the inspection procedure.

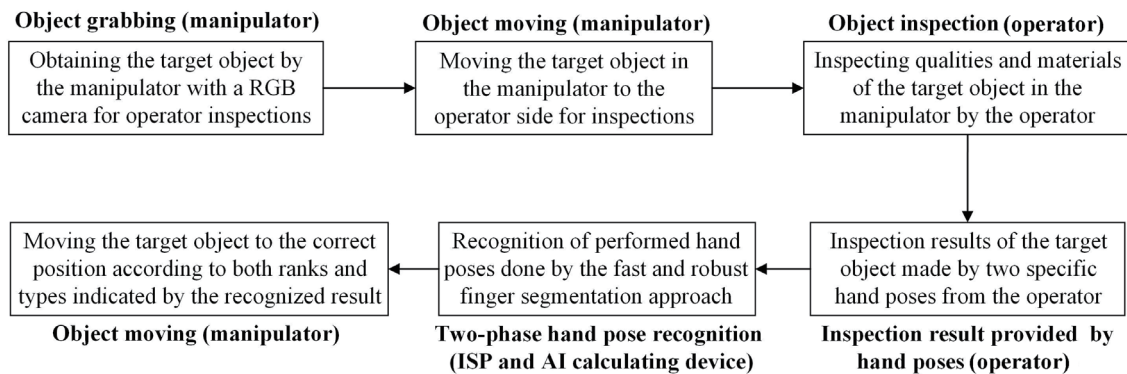


Fig. 3. Flowchart of the presented human-machine collaboration-based object inspection scheme including simultaneous operating actions of both the manipulator and the operator.

Step 4 (hand poses made by the operator, *operator* operations): After object inspection by the operator, the operator will perform hand poses to indicate the inspection evaluation results of the object (mainly one hand pose made for the object quality rank and the other hand pose made for the object material type).

Step 5 (hand pose recognition, *computing device* operations): This step includes the operations of hand poses of the operator. The main work of this step is to perform hand pose recognition, including image signal processing (ISP) with the noise removal of the skin-color regions of non-operator hands and the hand pose recognition of two phases, namely, one phase for the classification of numeric symbol hand pose images and the other phase for the classification of non-numeric expression hand pose images.

Step 6 (object moving from operator positions to pose indication positions, *manipulator* operations): This step is the final procedure and consists of the completions of hand pose recognition calculations. According to the indication revealed by the recognition result of two-phase hand pose recognition, this inspection object grabbed by the manipulator will finally be moved to an appropriate position and then released (usually the area of completed inspections in the work platform or in the automatic guide vehicle with the ability of carrying the object).

In the developed system of human manipulator collaboration-based object inspections with operator hand pose recognition, the manipulator adopted is made by the company ROBOTIS, which is the robot operating system (ROS)-enabled “OpenMANIPULATOR-X RM-X52-TNM.” This manipulator device is an open robot platform that finely supports “OpenHardware” to allow the user to adjust the robot device (e.g., modifications of link length) and “OpenCR,” an open-source control module for ROS, to enable the user to quickly develop the desired control strategy.⁽¹⁶⁾ In this work, U2D2 that belongs to a type of USB communication converter is used to conduct real-time communication between the computer and the manipulator (see Fig. 2, U2D2 connected to the USB port of the computer by the USB cable). The camera deployed in the robot device of OpenMANIPULATOR-X is a general web camera with the resolution of “Full HD-1080p.” Moreover, note that the camera is appropriately deployed above the robot gripper to

accurately capture the hand pose image of the operator if the manipulator moves the grabbed object to the front side of the operator.

The developed human manipulator collaboration-based object inspection system, consisting of both the hand pose image recognition system and the manipulator kinematic control system, is essentially a system integration structure that appropriately integrates two different types of operating system, namely, ROS and Windows. The recognition of the hand pose image of the operator is performed in the Windows platform. The kinematic control of the manipulator is carried out in the ROS platform. A virtual machine scheme (e.g., the use of VMware or Virtual Box) can effectively achieve such an integration of dual-system projects.

3. Robust and Fast Recognition of Hand Pose Images of Numeric Symbols Combined with Non-numeric Expressions for HRC-based Object Inspection

In this section, we will describe in detail the robust and fast approach for the classification of the hand pose images of numeric symbols combined with non-numeric expressions in this work (the procedure of Step 5 in the overall presented HRC-based object inspection scheme; also see Fig. 3 in the previous section). For online objection inspection with the requirement of real-time computations in the field application of smart factories, the fast finger segmentation approach proposed earlier in Ref. 15 and categorized into feature-based image pattern recognition types was extremely appropriate as the classifier of the hand poses of the operator in this work because of the neglect of the time-consuming model training and the avoidance of the non-economic pattern matching in the established models. Although the finger segmentation approach presented in Ref. 15 is competitive in terms of computation speed, in the practical application scenario of object inspection in a smart factory, two kernel technique issues will be inevitably encountered and require fine considerations: (1) frequent appearances of the noise data of the skin-color regions of non-operator hands in the captured scene image and (2) two different categorizations of hand poses performed by the operator, namely, the finger segmentation set of numeric symbols (denoting the object material type) and the finger segmentation set of non-numeric expressions (representing the object quality rank), which was compared with the classification of only finger segmentation sets with numeric symbols in Ref. 15. For tackling the above-mentioned issues and enhancing the typical finger segmentation hand pose recognition for further use in HRC-based object inspection in a smart factory, a robust design with effective noise removal (an ISP improvement on finger segmentation-based recognition) and two-phase hand pose recognition with the ability of clever separations between the finger segmentation sets of the numeric symbol type and those of the non-numeric expression type (a classification improvement on finger segmentation-based recognition) were developed in this work (see Sects. 3.2 and 3.3). The fast finger segmentation approach to the recognition of numeric symbol hand pose images presented in Ref. 15 will be primarily introduced in Sect. 3.1.

3.1 Fast numeric symbol hand pose recognition by finger segmentation

The main work of the finger segmentation-based approach to hand pose recognition earlier proposed in Ref. 15 is to determine the total number of finger segments in an acquired hand pose

image [also see Fig. 4]. When the presented method is used for the classification in a numeric set of 0, 1, 2, ..., 9, to obtain clear classifications in the case of the number of hand poses with the same finger segment number encountered (e.g., one finger segment of numbers 1 and 6, two finger segments of numbers 2 and 7, three finger segments of numbers 3 and 8, and four finger segments of numbers 4 and 9), a thumb detection algorithm was also developed in the work of Chen *et al.*⁽¹⁵⁾ Upon verifying whether a thumb segment exists, the problem of the separation of the numeric hand poses with the same finger segment number can be solved. For separating the part of only finger segments from a color camera-captured image (with the type of RGB), relative image process techniques including a binary-valued process (a conversion of RGB pixels to binary-valued pixels), the detection of the hand contour (to determine the hand region), the distance transform procedure (to determine the hand center), and wrist-line detection and the estimation of boundary points by sampling points in the constructed circle [developed in Ref. 15, to separate the part containing only finger segments from the hand pose image, also see Figs. 5(a) and 5(b)] are carried out in this order. For another issue of the verification of the thumb segment existence, a thumb detection approach by evaluating both the distance and angle of the hand center to the center of each separated finger segment is presented in Ref. 15 [also see Figs. 5(c) and 6].



Fig. 4. (Color online) Ten numeric hand poses, 0, 1, 2, ..., and 9, to be classified by the finger segmentation-based approach in Ref. 15.

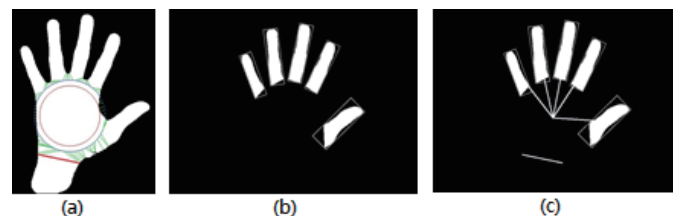


Fig. 5. (Color online) (a) Illustrations of the use of wrist-line detection and the estimate of boundary points by sampling points in the circle for separating the part of finger segments; (b) processed image with only the part of separated finger segments; and (c) thumb detection by evaluating the distance and angle of the hand center to the center of each finger segment.

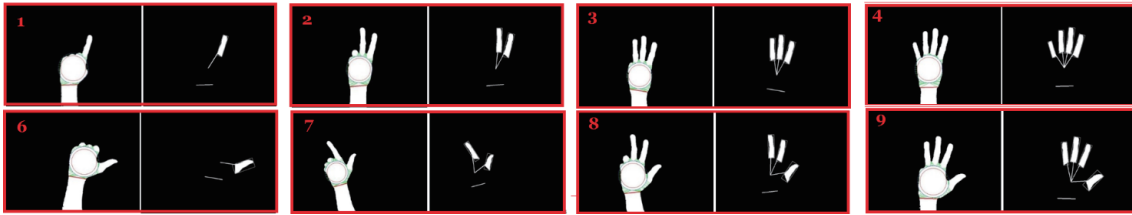


Fig. 6. (Color online) Illustrations of a clear separation of hand pose number sets of 1 and 6, 2 and 7, 3 and 8, and 4 and 9 by checking for the existence of a thumb segment by thumb detection.

Although the above-mentioned finger segment approach to hand pose recognition does not require a procedure for the establishment of classification models (i.e., the exhausting modeltraining procedure) and has the advantage of fast computations on recognition, two related kernel problems have yet to be solved satisfactorily. The first problem is the annoying phenomenon that the finger segment part has the noise data of the non-hand regions with the skin color; the second problem is an insufficient design of the classification of only the numeric hand poses available, which will be apparently insufficient in the case that such finger segmentation-based hand pose recognition is used in a practical human–robot application scenario (such as the HRC-based object inspection application scenario in a smart factory as mentioned in Sect. 2). Aimed at these required techniques for enhancing finger segmentation-based hand pose recognition, the problems of the complex scene with a noisy background and the restricted classification labels with only numeric hand poses will be thoroughly considered, and the corresponding improvements are discussed in Sects. 3.2 and 3.3, respectively (see Fig. 7).

3.2 Robust finger segmentation-based hand pose recognition with removal of noisy data of the skin-like region that is not part of the operator’s hand

In the case that finger segmentation-based hand pose recognition is performed in a clean environment and that the captured scene image is simple and contains only one skin-color region of the operator hand, typical finger segmentation-based hand pose recognition can perform well by only focusing on the found hand region for further analysis by the approach mentioned in Sect. 3.1 [see Fig. 8(a)]. However, in practical real-life applications such as the smart factory application of human–machine collaboration-based object inspection with attendances of the product-line operator, a complex environment will be inevitably encountered, in which a complex scene image including various types of skin-color region, the operator hand region, non-hand regions with skin colors (such as faces, necks, arms, and backgrounds), and the non-operator hand region (persons that appear behind the operator and make hand poses), will be obtained [see Fig. 8(b)]. In such a case, the typical finger segmentation-based approach will be extremely difficult to use for hand pose recognition as the binary-valued image from the captured scene contains the target operator hand region and various other significant regions with the skin color to be analyzed. For solving this problem, we will need to remove the noisy data of the skin-color regions of non-operator hands from this evaluated binary-valued scene image, as presented in detail in the following text.

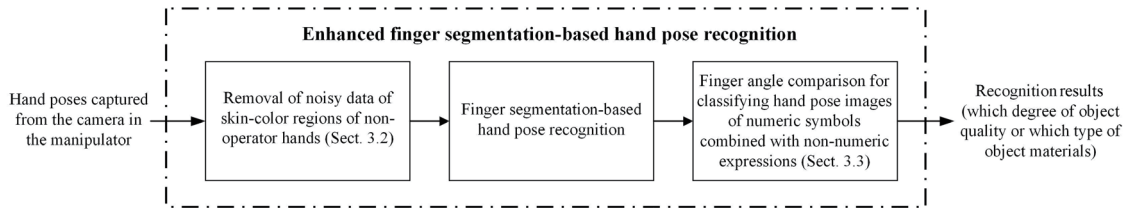


Fig. 7. Flowchart of enhanced finger segmentation-based hand pose recognition.



Fig. 8. (Color online) Illustrations of two different scenes containing the operator hand region: (a) simple and (b) complex scene.

In this study, for thoroughly removing the noisy data that were not the operator hand region, two ways were considered: (1) hand contour analysis on all of the obtained skin-color regions and (2) operator hand estimates on all of the acquired hand regions. For the hand contour analysis, three geometric features were used to detect whether all of the found contours containing the skin color were the hand region. These three geometric features were solidity, circularity, and extent, which can be calculated using Eqs. (1)–(3), respectively, as follows:

$$Solidity = \frac{contour_area}{convex_hull_area}, \quad (1)$$

$$Circularity = \frac{4\pi \cdot contour_area}{contour_perimeter^2}, \quad (2)$$

$$Extent = \frac{contour_area}{rectangle_border_area}, \quad (3)$$

where the four parameters *counter_area*, *convex_hull_area*, *contour_area* and *rectangle border_area* represent the different categorizations of the geometric characteristics revealed by the estimated skin-color region, the calculated area of the contour [see Fig. 9(a)], the calculated area of the convex hull [see Fig. 9(b)], the calculated perimeter of the contour [see Fig. 9(a)], and the calculated area of the rectangle border [see Fig. 9(c)]. Regions with the skin color in the captured scene image were categorized into the hand type only when all of the computed values of *Solidity*, *Circularity*, and *Extent* fell into valid ranges, which were set as [0.6, 0.75], [0.1, 0.25], and [0.4, 0.6], respectively, in this work.

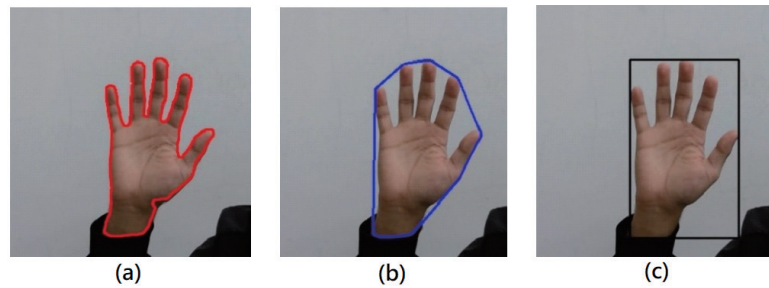


Fig. 9. (Color online) Different categorizations of geometric characteristics used for acquiring the target hand region among all of the detected regions with the skin color in the capture scene: (a) contour area and perimeter; (b) convex hull area; and (c) rectangle border area.

Followed by the above-mentioned hand contour analysis on all of the obtained skin-color regions (non-hand skin-color regions removed and only hand regions of the operator or the non-operator kept in the binary-valued image), the operator hand estimate on all of the acquired hand regions was then carried out. In this study, such an estimate to determine which hand region belonged to the operator was designed in a simple and fast manner. All of the pixels around the hand contour of the found hand region were accumulated. Among all of the found hand contours to be evaluated, the hand contour of the operator apparently had the maximum number of accumulated pixels, by which the hand belonging to the operator can be determined. Figure 10 shows such criteria for the determination of the target operator hand.

3.3 Two-phase recognition of numeric and non-numeric hand pose images for identification of the object type and categorization of the object quality

Figure 11 shows two different usage categorizations of hand poses, namely, the numeric and non-numeric expression types of hand pose, to be performed by the operator in the work of HRC-based object inspection. As shown in Fig. 11, nine hand poses categorized into numeric symbols, 1, 2, ..., 9, represent the corresponding nine different object material types; in the non-numeric expression type of hand pose, five hand poses were designed in this work to indicate the quality rank of the inspected object, which were Rank-A (RA), Rank-B (RB), Rank-C (RC), Rank-D (RD), and Rank-E (RE) in the order of the quality distributed from the best to the worst. As mentioned in Sect. 2 (also see Fig. 3), after the completion of the task of object inspection by the operator in Step 3, the main work in Step 4 was that, according to the inspected result obtained by the operator, the operator performed two different hand poses, namely, one pose denoting the object quality and the other pose denoting the object material, to tell the system to which position the object grabbed by the manipulator had to be moved and released. The typical finger segmentation approach as mentioned in Sect. 3.1 did not work well for the recognition of these designed 14 hand poses containing simultaneously both the numeric and non-numeric hand poses (difficulty occurring in the recognition of poses RB and 8, poses RC and 2, poses RD and 3, poses RE and 9, and poses 6 and 7). In the cases of poses RB and 8, poses RE and 9, and poses 6 and 7, after the computation of the typical finger segmentation approach, the

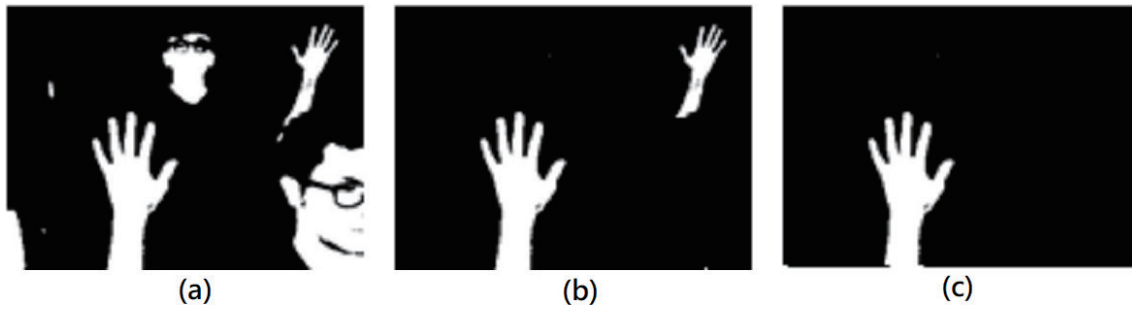


Fig. 10. Removal of noisy data of skin-color regions of non-operator hands: (a) original binary-valued scene image without noisy data removal; (b) image obtained after hand contour analysis; (c) image obtained after hand contour analysis and operator hand estimates.

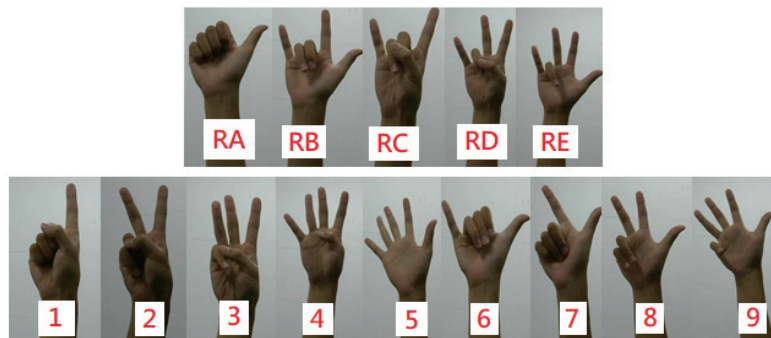


Fig. 11. (Color online) Hand poses of nine numeric symbols (denoting object material types, 1, 2, ..., 9) combined with five non-numeric expressions (denoting object quality ranks, RA, RB, RC, RD, and RE) designed for human manipulator collaboration-based object inspection in a smart factory.

recognition results of three finger segments with the thumb, four finger segments with the thumb, and two finger segments with the thumb, respectively, were finally obtained (see Fig. 12). For these two situations of poses RC and 2 and poses RD and 3, with typical finger segmentation, the recognition outcomes of two and three finger segments without the thumb were obtained, respectively (see Fig. 12). In all of the above-mentioned confusing hand pose sets, the main problem of typical finger segmentation-based hand pose recognition was the recognition result with “the same number of finger segments” and “the same result on the appearance of the thumb” for each confusing hand pose set. To overcome this problem and for a clear separation in each of the confusing hand pose sets, namely, (RB, 8), (RC, 2), (RD, 3), (RE, 9), and (6, 7), a finger angle approach algorithm is proposed herein to further enhance the typical finger segmentation-based hand pose recognition. Such a presented finger angle comparison approach can be appropriately incorporated with the typical finger segmentation approach to further achieving the ability of classifying hand poses containing both numeric and non-numeric types.

The procedure for the presented finger angle comparisons is as follows.

Step 1: According to the binary-valued image of hand poses derived from fast and robust finger

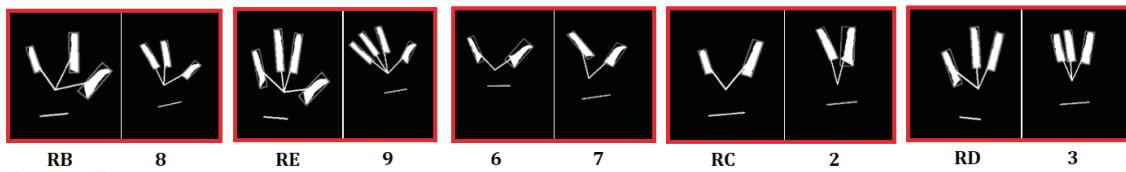


Fig. 12. (Color online) Illustrations of confusing hand pose sets, (RB, 8), (RE, 9), (6, 7), (RC, 2), and (RD, 3), required to be further classified by the presented finger angle comparison method.

segmentation calculations, the total number of finger segments is first estimated.

Step 2: Among all of the found finger segments, for all of the possible conditions of two arbitrarily chosen finger segments, the segment angles are computed.

Step 3: In the set of segment angles acquired in Step 2, the maximum segment angle (to be denoted as θ_{max}) is chosen for further verification of the confusing hand poses.

Step 4: Two confusing hand poses can be clearly separated according to θ_{max} estimated in Step 3. According to the current condition of the finger segment number and thumb appearance, the corresponding threshold of angles is selected from the table of maximum segment angle thresholds of confusing hand pose sets (see Table 1, which is established in advance) for making comparisons with θ_{max} .

For a detailed explanation of the usage of the above-mentioned finger angle comparison approach, an example of the classification of the confusing hand poses RE and 9 is provided (see Fig. 13, images acquired after robust and fast finger segmentations). As depicted in Fig. 13, both of the poses of RE and 9 had the same condition of four finger segments and thumb appearance. Four finger segments were numbered 1, 2, 3, and 4 (as shown in Fig. 13), and in all, six finger segment angles were calculated, namely, angles of (1, 2), (1, 3), (1, 4), (2, 3), (2, 4), and (3, 4). Among all of the six calculated angles, the maximum angle, angle of (1, 4), was then derived and denoted as θ_{max} . θ_{max} was then evaluated according to Table 1 (mainly the second condition of the four segments and thumb detected in Table 1). According to Table 1, if θ_{max} was larger than 120° , the confusing hand pose was classified as RE; otherwise, the classification label of the pose was determined to be 9. Similar to the classifications of poses RE and 9, the separations of the confusing hand poses of the other four conditions could also be effectively carried out. Note that as shown in Table 1, different threshold values of 100° , 90° , 50° , and 60° were set for the separations of poses RB and 8 (1st condition), poses 6 and 7 (3rd condition), poses RC and 2 (4th condition), and poses RD and 3 (5th condition), respectively. Note that each of these threshold values, namely, 120° , 100° , 90° , 50° , and 60° listed in Table 1 for a clear separation of the corresponding confusing hand pose set with the same finger number was determined by using only a simple hand shape geometry approach.

With the appropriate incorporation of the presented finger angle comparisons, robust and fast finger segmentation-based hand pose recognition was then further enhanced for a significant extension from the typical classifications of only simple numeric symbols to the advanced recognition of complex hand pose labels containing simultaneously both numeric symbols and non-numeric expressions. Such two-phase hand pose recognition wherein the operator made the non-numeric expression pose and the numeric symbol pose to represent the object quality rank

Table 1

Comparison of maximum segment angle thresholds of confusing hand pose sets.

Conditions	Threshold (TH) ($^{\circ}$)	$\theta_{max} > TH$	$\theta_{max} \leq TH$
1 (three segments, thumb detected)	100	Pose RB	Pose 8
2 (four segments, thumb detected)	120	Pose RE	Pose 9
3 (two segments, thumb detected)	90	Pose 6	Pose 7
4 (two segments, no thumb)	50	Pose RC	Pose 2
5 (three segments, no thumb)	60	Pose RD	Pose 3

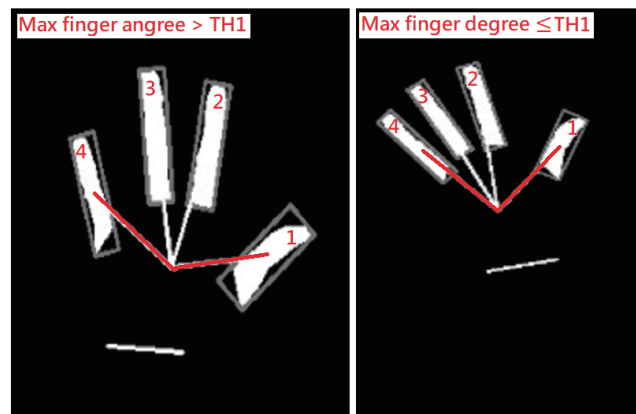


Fig. 13. (Color online) Example of two confusing hand poses with same conditions of four finger segments and thumb appearance, RE (left) and 9 (right), to be classified by finger angle comparisons.

(the 1st phase recognition for classifying poses RA, RB, RC, RD, and RE) and then the object material type (the 2nd phase recognition for classifying poses 1, 2, ..., and 9), respectively, was effectively applied to a smart factory with human manipulator collaboration-based object inspections. An example of the use of the two-phase hand pose calculation is that the object inspection result “RA-1” denotes that the object has quality rank A and material type 1. In all, there were nine object inspection results that contained both the quality rank and the material type designed in the two-phase recognition, which will be discussed in detail in Experiments (Sect. 4).

4. Experiments

Figure 14(a) depicts the operation environment of the presented human manipulator inspection deployed in a laboratory office, which was a scenario simulation of the practical smart factory. Figure 14(b) shows the manipulator with a well-deployed camera used as an assistant in an HRC inspection task as mentioned previously. As shown in Fig. 14(a), the environment was established to have the uninspected area, the inspection finished area, a personal computer (PC) for performing edge AI computations of the recognition of the operator’s hand poses, a panel screen to show the hand pose recognition results, and the manipulator with a camera [also see Fig. 14(b)]. In the situation that the HRC inspection task was not started, the

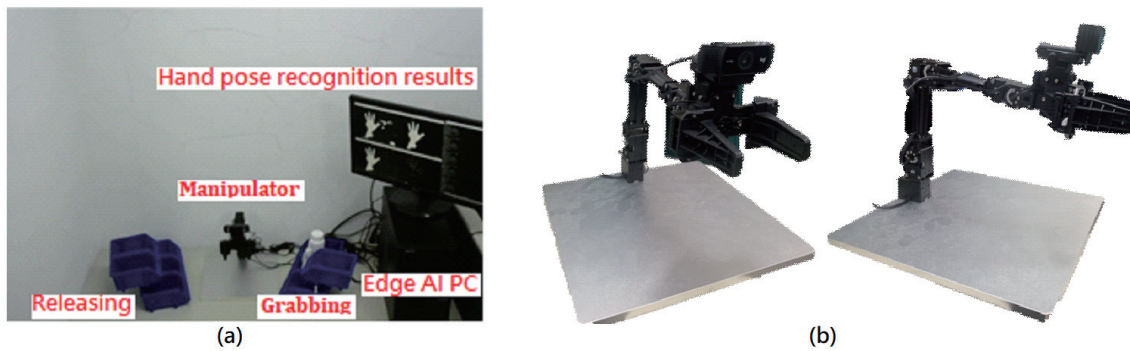


Fig. 14. (Color online) Human-machine collaboration-object inspection implemented in real-life applications (a) in a laboratory office environment and (b) the manipulator with a general camera.

manipulator would be in the ready state and kept motionless. When obtaining the command to begin the work of object inspections, the manipulator would perform a series of actions to help the operator to efficiently complete the object inspection mission, which included mainly moving to the uninspected area, grabbing the object in the uninspected area, moving to the initial ready position for the operator's inspection, waiting for operator inspection (object inspection performed by eye-seeing, hand-touching, nose-smelling, and expert domain knowledge of the operator), waiting for the inspection result obtained (including mainly quality ranks and material types represented by two consecutive hand poses of the operator), waiting for hand pose recognition (as described in Sects. 3.1–3.3) performed in the PC, moving to the corresponding position in the inspection-completed area according to the hand pose recognition result, releasing the grabbed object in the manipulator, and finally returning to the initial position in order to be ready for performing the next object inspection task. Such detailed implementations of the presented human manipulator cooperative-object inspection scheme incorporated with the two-phase hand pose recognition are shown in Fig. 15. Figure 16 shows the detailed process outcome of the presented two-phase hand pose recognition shown in Fig. 15 where the recognition calculations of the two consecutive hand poses of the operator for obtaining the object inspection results, RA (denoting the quality rank) and 1 (denoting the material type), were carried out, and the recognition result RA-1 was thus obtained.

Tables 2 and 3 present the databases of hand poses to denote the corresponding five different object quality ranks (labeled RA, RB, RC, RD, and RE) and of hand poses to represent the corresponding nine different object material types (labeled 1, 2, ..., and 9), respectively. Note that in Tables 2 and 3, the initial hand pose images before segmentation were those that had undergone noise removal (using the presented approach discussed in Sect. 3.2). The hand pose images processed by finger segmentation (mentioned in Sect. 3.1) are also listed in these tables (the column of "after segmentation"). In addition, for each categorization of hand poses, the last column of these tables also provides what methods (or which method) could be adopted for a clear label separation in the situation that hand poses with the same number of finger segments were encountered. For hand pose 5, because of the uniqueness of the five finger segments in all

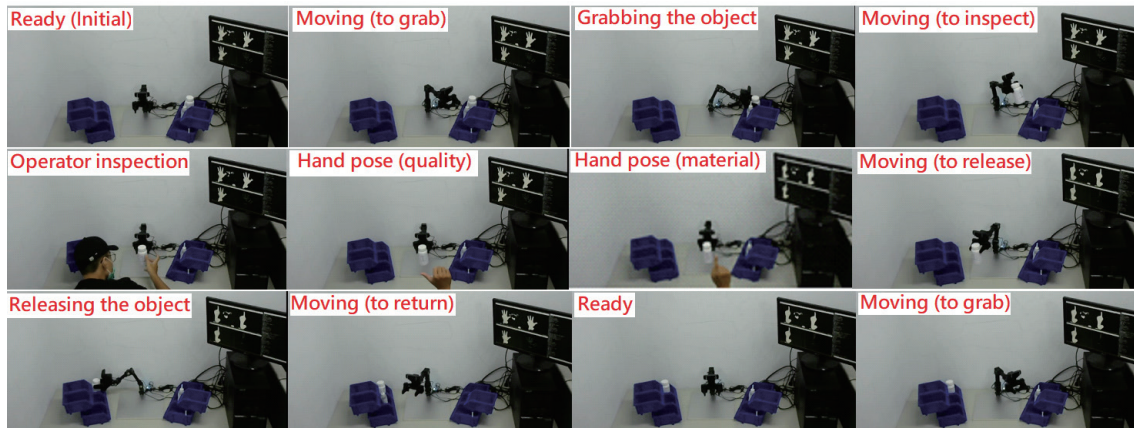


Fig. 15. (Color online) Real-life implementation of presented human–machine collaboration–object inspection incorporated with two-phase hand pose recognition (various operator manipulator-cooperation states shown, also see the system flowchart mentioned in Fig. 3).

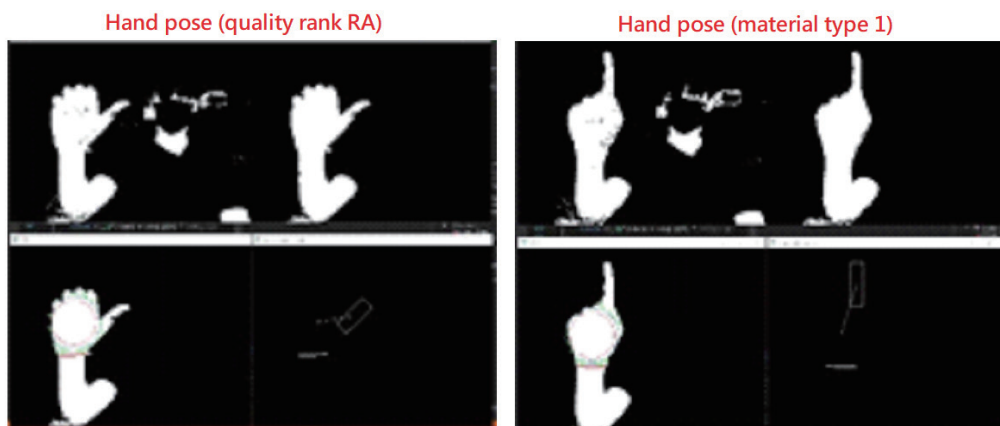


Fig. 16. (Color online) Two-phase hand pose recognition (two consecutive hand poses, RA and 1, performed by the operator) to indicate the inspection result; hand pose (quality) and hand pose (material) in Fig. 15.

of the 14 hand poses, only the finger segmentation approach was required. In these cases of “hand poses with one finger segment, namely, RA and 1,” “hand poses with two finger segments, namely, RC, 2, 6, and 7,” “hand poses with three finger segments, namely, RB, RD, 3, and 8,” “hand poses with four finger segments, namely, RE, 4, and 9,” both finger segmentation and thumb detection approaches were required. In the case of hand poses with two finger segments, when the thumb was detected, the possible hand pose would be 6 or 7; otherwise, the hand pose was RC or 2. In the case of hand poses with three finger segments, when the thumb appeared, the possible hand pose was RB or 8; when no thumb was detected, the hand pose was RD or 3. In the hand poses with four finger segments, when the thumb was not detected, the hand pose was 4; when the thumb was observed, RE or 9 was the possible hand pose. As mentioned earlier, for the confusing hand pose sets, namely, (RB, 8), (RE, 9), (6, 7), (RC, 2), and (RD, 3), the finger angle comparisons described in Sect. 3.3 were additionally required for the separation between two confusing poses of the set.

Table 2

(Color online) Database of five different hand poses to denote corresponding object quality ranks, RA, RB, RC, RD, and RE (before segmentation, after segmentation, and criterion used on recognition).

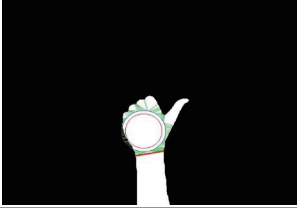
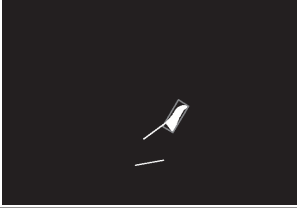
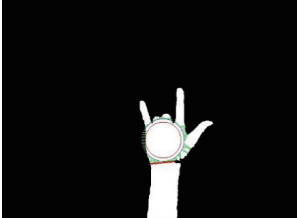
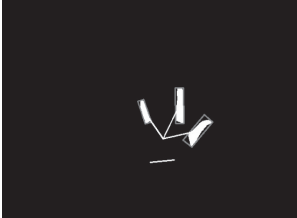
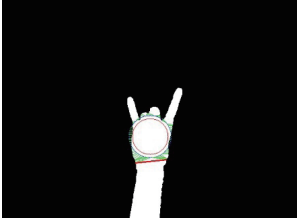

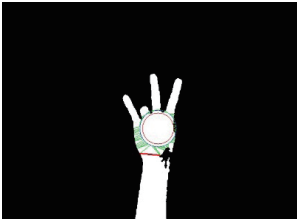

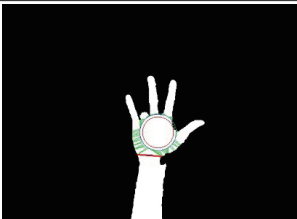

Rank label (quality type)	Before segmentation	After segmentation	Criterion used to process the same number of finger segments estimated
RA			<ul style="list-style-type: none"> • 1 finger segment • Thumb detected • To be distinguished from hand pose 1
RB			<ul style="list-style-type: none"> • 3 finger segments • Thumb detected • To be distinguished from hand poses RD and 3 • To be separated from hand pose 8 by finger angle comparisons
RC			<ul style="list-style-type: none"> • 2 finger segments • No thumb • To be distinguished from hand poses 6 and 7 • To be separated from hand pose 2 by finger angle comparisons
RD			<ul style="list-style-type: none"> • 3 finger segments • No thumb • To be distinguished from hand poses RB and 8 • To be separated from hand pose 3 by finger angle comparisons
RE			<ul style="list-style-type: none"> • 4 finger segments • Thumb detected • To be distinguished from hand pose 4 • To be separated from hand pose 9 by finger angle comparisons

Table 4 shows the performance of the presented removals of the noisy data of the skin-color regions of non-operator hands (see Sect. 3.2). As shown in Table 4, four background environments (Cases 1 to 4), each of which contained different degrees of noisy data (from slight to serious), were evaluated. It was clearly observed that the presented noise removal approach in Sect. 3.2 could perform effectively in all of these cases, achieving the ability of thoroughly eliminating all of the skin-color regions of non-operator hands (i.e., only the hand region of the operator was preserved).

Table 3

(Color online) Database of nine different hand poses to denote corresponding object material types, 1, 2, ..., 9 (before segmentation, after segmentation, and criterion used on recognition).

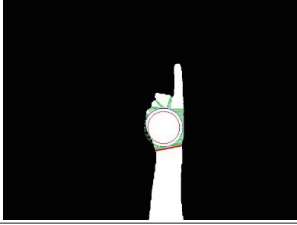
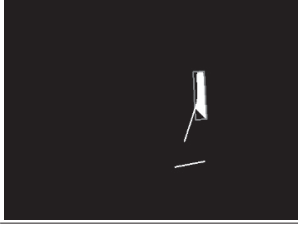

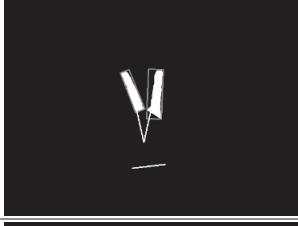
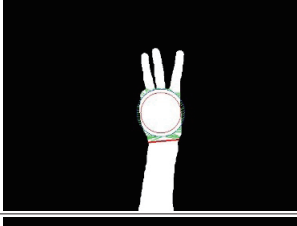
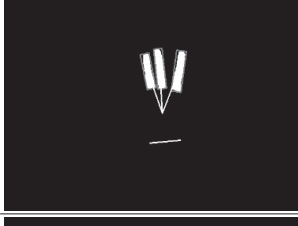
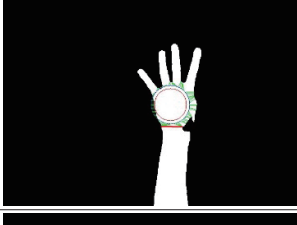
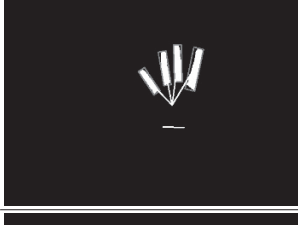
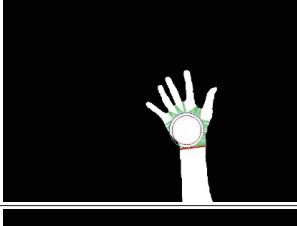
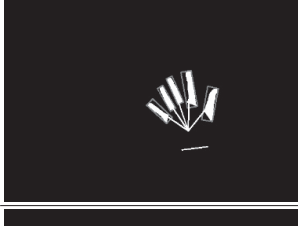
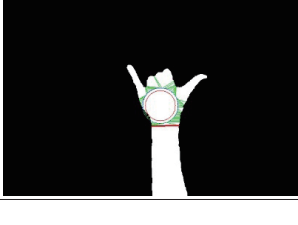
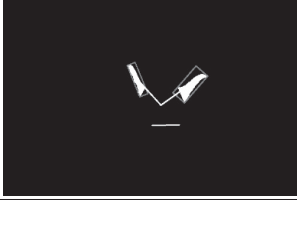
Product label (material type)	Before segmentation	After segmentation	Criterion used to process the same number of finger segments estimated
1			<ul style="list-style-type: none"> • 1 finger segment • Thumb detected • To be distinguished from hand pose RA
2			<ul style="list-style-type: none"> • 2 finger segments • No thumb • To be distinguished from hand poses 6 and 7 • To be separated from hand pose RC by finger angle comparisons
3			<ul style="list-style-type: none"> • 3 finger segments • No thumb • To be distinguished from hand poses RB and 8 • To be separated from hand pose RD by finger angle comparisons
4			<ul style="list-style-type: none"> • 4 finger segment • No thumb • To be distinguished from hand poses RE and 9
5			<ul style="list-style-type: none"> • 5 finger segments
6			<ul style="list-style-type: none"> • 2 finger segments • Thumb detected • To be distinguished from hand poses RC and 2 • To be separated from hand pose 7 by finger angle comparisons

Table 3

(Continued) (Color online) Database of nine different hand poses to denote corresponding object material types, 1, 2, ..., 9 (before segmentation, after segmentation, and criterion used on recognition).

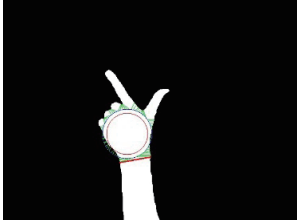
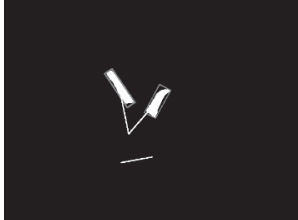
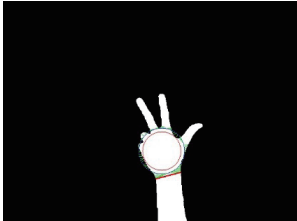
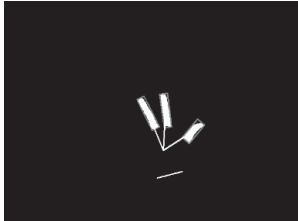
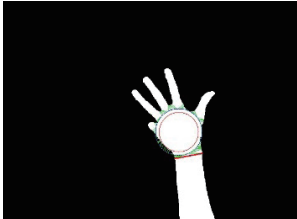
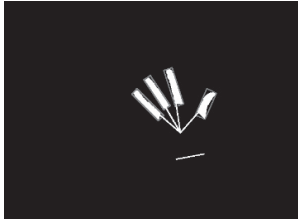
Product label (Material type)	Before segmentation	After segmentation	Criterion used to process the same number of finger segments estimated
7			<ul style="list-style-type: none"> • 2 finger segments • Thumb detected • To be distinguished from hand poses RC and 2 • To be separated from hand pose 6 by finger angle comparisons
8			<ul style="list-style-type: none"> • 3 finger segments • Thumb detected • To be distinguished from hand poses RD and 3 • To be separated from hand pose RB by finger angle comparisons
9			<ul style="list-style-type: none"> • 4 finger segments • Thumb detected • To be distinguished from hand pose 4 • To be separated from hand pose RE by finger angle comparisons

Table 5 shows the confusion matrix and the recognition accuracy of the presented robust and fast hand pose recognition approach with the collected hand pose database containing 14 different classes of hand poses (as mentioned, object quality ranks, namely, RA, RB, RC, RD and RE, and object material types, namely, 1, 2, ..., and 9). Note that for each same category of hand poses, ten times of poses are collected. In all, 140 hand pose images were included in the database. As observed from Table 5, the recognition accuracy achieved 100% of complete recognition, which would be an extremely satisfactory performance for the use of human manipulator collaborative-object inspections in a practical smart factory scenario. Note that the performance results of the classifications of all of the 14 hand pose types in Table 5 revealed that in a practical online object inspection task of two-phase hand pose recognition, two consecutive hand poses were classified; nine mixed hand pose inspection sets, namely, RA-1, RA-2, RB-3, RB-4, RC-5, RC-6, RD-7, RD-8, and RE-9, could also achieve the perfect classification accuracy of reaching complete recognition. Tables 6 and 7 present the calculation speed performances of the presented hand pose recognition in the classifications of one hand pose and two consecutive hand poses in the two-phase recognition, respectively (PC with Windows 10, CPU of Intel i5

Table 4

(Color online) Removals of noisy data of skin-color regions of non-operator hands performed before finger segmentations (four different degrees of noisy conditions, from slight to serious).




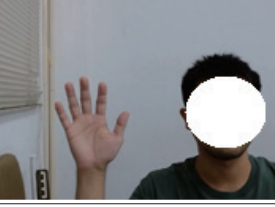








Cases of noises	Original RGB images without hand detection	Detection of hands of the operator without noise removal	Detection of hands of the operator with noise removal
1			
2			
3			
4			

Table 5

Confusion matrix and recognition accuracy of the established hand pose database of fourteen different hand poses (five poses for quality ranks and nine poses for material types) using fast and robust finger segmentation-based hand pose segmentation with finger angle comparisons.

	1	2	3	4	5	6	7	8	9	RA	RB	RC	RD	RE
1	10	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	10	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	10	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	10	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	10	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	10	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	10	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	10	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	10	0	0	0	0	0
RA	0	0	0	0	0	0	0	0	0	10	0	0	0	0
RB	0	0	0	0	0	0	0	0	0	0	10	0	0	0
RC	0	0	0	0	0	0	0	0	0	0	0	10	0	0
RD	0	0	0	0	0	0	0	0	0	0	0	0	10	0
RE	0	0	0	0	0	0	0	0	0	0	0	0	0	10

Average recognition accuracy: 100%

3.40 GHz, and RAM of 8 GB). As observed from Table 6, the average recognition time to recognize a hand pose was only 1.81 s. In the task of human manipulator-collaborative object inspections, the two-phase hand pose recognition with two consecutive hand poses to be classified would still remain competitive in the average calculation time, achieving an acceptable performance of only 3.67 s. The experimental results of hand pose recognition with superior recognition accuracy and economic calculation cost revealed that the human manipulator-collaborative object inspection would be a considerably feasible method to be realized in the practical smart factory.

Table 6

Calculation speed performance of the established hand pose database of fourteen different hand poses using fast and robust finger segmentation-based hand pose segmentation with finger angle comparisons (one-phase recognition, only one hand pose recognized).

	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Avg. (sec.)
1	2	1	1	1	1	1	1	1	2	1	1.2
2	2	1	2	2	2	2	2	2	2	1	1.8
3	2	2	2	2	2	2	2	2	2	2	2
4	2	2	2	2	2	2	2	2	2	2	2
5	2	2	2	2	2	2	2	2	2	2	2
6	2	2	1	1	2	1	1	1	2	2	1.5
7	2	1	2	2	2	2	2	2	2	2	1.9
8	2	2	2	2	2	2	2	2	2	2	2
9	2	2	2	2	2	2	2	2	2	2	2
RA	1	1	1	1	1	1	1	1	1	1	1
RB	2	2	2	2	2	2	2	2	2	2	2
RC	2	2	2	2	2	2	2	2	2	2	2
RD	2	2	2	2	2	2	2	2	2	2	2
RE	2	2	2	2	2	2	2	2	2	2	2

Average recognition time: 1.81.

Table 7

Calculation speed performance of the established hand pose database of fourteen different hand poses using fast and robust finger segmentation-based hand pose segmentation with finger angle comparisons (two-phase recognition, two consecutive hand poses recognized).

	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	Avg. (sec.)
RA-1	3	3	3	3	3	3	3	3	3	3	3
RA-2	3	3	3	3	3	3	3	3	3	3	3
RB-3	4	4	4	4	4	4	4	4	4	4	4
RB-4	4	4	4	4	4	4	4	4	4	4	4
RC-5	4	4	4	4	4	4	4	4	4	4	4
RC-6	3	3	3	3	3	3	3	3	3	3	3
RD-7	4	4	4	4	4	4	4	4	4	4	4
RD-8	4	4	4	4	4	4	4	4	4	4	4
RE-9	4	4	4	4	4	4	4	4	4	4	4

Average recognition time: 3.67.

5. Conclusions

In this work, a human manipulator collaboration-based object inspection scheme was proposed, and a robust and fast two-phase hand pose recognition approach to recognizing two consecutive hand poses (to denote the object quality rank and object material type) of the operator was developed. Experiments in the categorizations of 14 different hand poses, including five hand pose classes of non-numeric expressions for object quality ranks and nine hand pose classes of numeric symbols for object material types, demonstrate the effectiveness and efficiency of the presented hand pose recognition. The developed human manipulator collaboration-based object inspection incorporated with the hand pose recognition of the operator in this study is considerably competitive and can further promote current factories to be more intelligent.

Acknowledgments

We acknowledge the support given by the National Science and Technology Council (NSTC), Taiwan under Grant no. 111-2221-E-150-034.

References

- 1 Z. Huang, Y. Shen, J. Li, M. Fey, and C. Brecher: *Sensors* **21** (2021) 6340. <https://doi.org/10.3390/s21196340>
- 2 A. Md. Hazrat, K. Aizat, K. Yerkhan, T. Zhandos, and O. Anuar: *Procedia Comput. Sci.* **133** (2018) 205. <https://doi.org/10.1016/j.procs.2018.07.025>
- 3 R. Yan, L. Jackson, and S. Dunnett: *J. Manuf. Syst.* **57** (2020) 19. <http://dx.doi.org/10.1016/j.jmsy.2020.08.005>
- 4 G. Fragapane, R. Koster, F. Sgarbossa, and J. O. Strandhagen: *Eur. J. Oper. Res.* **294** (2021) 405. <https://doi.org/10.1016/j.ejor.2021.01.019>
- 5 H.-C. Liao, Z.-Y. Lim, Y.-X. Hu, and H.-W. Tseng: *Proc. 2018 IEEE 3rd Int. Conf. Signal and Image Processing (IEEE, 2018)* 362–366. <https://doi.org/10.1109/SIPROCESS.2018.8600456>
- 6 V. Reshadat and R. A. J. W. Kapteijns: *Proc. 2021 Int. Conf. Data and Software Engineering (IEEE, 2021)* 1–5. <https://doi.org/10.1109/ICoDSE53690.2021.9648445>
- 7 Y. H. Tsai, N. Y. Lyu, S. Y. Jung, K. H. Chang, J. Y. Chang, and C. T. Sun: *Proc. 2019 IEEE/ASME Int. Conf. Advanced Intelligent Mechatronics (IEEE, 2019)* 103–107. <https://doi.org/10.1109/AIM.2019.8868602>
- 8 C.-Y. Huang, I.-C. Lin, and Y.-L. Liu: *Appl. Sci.* **12** (2022) 2269. <https://doi.org/10.3390/app12052269>
- 9 R. Vaga and K. Bryant: *Proc. 2019 Pan Pacific Microelectronics Symposium (IEEE, 2019)* 1–3. <https://doi.org/10.23919/PanPacific.2019.8696655>
- 10 R. Ye, M. Chang, C.-S. Pan, C. A. Chiang, and J. L. Gabayno: *Int. J. Optomechatronics* **12** (2018) 1. <https://doi.org/10.1080/15599612.2018.1444829>
- 11 I. J. Ding and J. L. Su: *Proc. Inst. Mech. Eng. Pt. B J. Eng. Manufact.* **237** (2022). <https://doi.org/10.1177/09544054221102247>
- 12 I. J. Ding and N. W. Zheng: *Sensors* **22** (2022) 803. <https://doi.org/10.3390/s22030803>
- 13 I. J. Ding, N. W. Zheng, and M. C. Hsieh: *J. Intell. Fuzzy Syst.* **40** (2021) 7775. <http://dx.doi.org/10.3233/JIFS-189598>
- 14 I. J. Ding and M. C. Hsieh: *Microsyst. Technol.* **28** (2022) 403. <https://doi.org/10.1007/s00542-020-04868-9>
- 15 Z.-H. Chen, J.-T. Kim, J. Liang, J. Zhang, and Y.-B. Yuan: *Sci. World J.* **2014** (2014) 267872. <https://doi.org/10.1155/2014/267872>
- 16 OpenMANIPULATOR-X. https://emmanual.robotis.com/docs/en/platform/openmanipulator_x/specification/ (retrieved in December 2022).