Wavelet Subband-Based Tensor for Smartphone Physical Button Inspection

DUONG BINH GIAP[1], **TUYEN NGOC LE**[2], **JING-WEIN WANG**[2], **(Member, IEEE), AND CHIA-NAN WANG**[3]

[1]Department of Electronics Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan
[2]Institute of Photonics Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan
[3]Department of Industrial Engineering and Management, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan

Corresponding authors: Tuyen Ngoc Le (tuyennl75@gmail.com) and Jing-Wein Wang (jwwang@nkust.edu.tw)

ABSTRACT A smartphone contains many critical components that are produced in highly automated and precisely monitored facilities throughout the complex manufacturing process. Even with the rapid development in the smartphone manufacturing industry today, the physical buttons are still existing on the smartphone because of the crucial importance of both in terms of their functionality and role. The smartphone's physical buttons are small in size and have non-planar and shiny surfaces that lead to difficulty in detecting defects not only with human eyes but also with most AOI systems. Besides, most defects are tiny, with low contrast which is a huge challenge for deep learning models-based defect detection. To overcome these challenges, we propose a novel framework based on machine vision named highlight defect region by using higher-order singular value decomposition of wavelet subband-based tensor (HHoWST) for real-time smartphone's physical buttons quality inspection. First, a modern image acquisition system is designed to obtain a high-quality smartphone's physical button image dataset with a total of 500 images containing 13,472 samples of six defect types. Next, a wavelet subband-based third-order tensor of the smartphone's physical button color image is constructed. Finally, higher-order singular value decomposition is proposed to estimate the components that contain the global illumination information and highlight the defective regions of the image. The experiments performed on HHoWST images reveal that our proposed method significantly improves the defect detection efficiency of deep learning models, such as SSD, Faster R-CNN, and YOLOv5, especially the performance in detecting the tiny defect types.

INDEX TERMS Automatic optical inspection, defect detection, higher-order singular value decomposition, tensor, wavelet transform, deep learning model.

I. INTRODUCTION

In recent years, the smartphone manufacturing industry has grown rapidly. A smartphone company not only depends on its own manufacturing alone but also has a lot of suppliers that it relies on for procuring components for assembly. Each component of a smartphone has a crucial role in ensuring its quality. Even though smartphone manufacturers have been clear that smartphone is being developed with the trend of all on-screen buttons. However, manufacturers continue putting physical buttons (hardware keys) in their devices, such as volume up/down button, side button, ring/silent switch, etc. These physical buttons have some disadvantages, such as taking more effort to utilize, be used frequently,

can't be changed, moved, or updated. Therefore, the most important factor of a physical button is durability, which leads to rigorous requirements in controlling the quality of the product with the components. Due to the small size and the non-planar surfaces of the smartphone's physical button, the defects existing on the surface of the physical button cannot be detected simply. Unfortunately, in most smartphone manufacturers, the visual inspection task is being conducted by human inspectors. The authors in [1] have shown that the human visual system declines with dull and endlessly routine jobs and cannot perform inspection in unfavorable environments. In order to improve the visual inspection efficiency, Automatic Optical Inspection (AOI) systems [2] are widely studied and used in different industrial fields, such as printed circuit board (PCB) [3]–[6], fabric [7]–[10], steel [11]–[14], glasses [15], [16], liquid crystal

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

displays (LCDs) [17]–[19], and many others. By using AOI, the inspection time can be reduced, while the accuracy during the inspection process is increased.

The AOI basic system contains three primary parts: image acquisition part, image preprocessing part, and defect inspection part. The image acquisition part includes cameras and illuminating lights to capture the product images. Then, the collected images are preprocessed to enhance the characterizing of defects, such as removing the noise and non-uniform backgrounds contained in the product images. Finally, the preprocessed images are used as input of the defect inspection part to detect and classify defects. In an AOI system, the image preprocessing part is vital to enhance the accuracy rate in detection and classification. Since a raw image acquired from the image acquisition system contains all kinds of information, some are favorable for the inspection process, and some are redundant. The favorable information is the feature of the defect used effectively to improve the judgment rate in the defect inspection step. On the other hand, redundant information is the leading cause in reducing the recognition rate of an AOI system. By using image processing algorithms embedded in the image preprocessing part of the AOI system, this favorable information (the defect feature) can be enhanced, whereas the redundant information is eliminated. It results in the improvement of the defect recognition performance of the AOI system. Various methods have been developed for image preprocessing, which can be categorized based on the domain in which the raw image is processed. In this paper, we group these methods into two primary categories: spatial domain methods and frequency domain methods.

The spatial domain methods involve many operations such as spatial filtering and geometric transformations used to enhance the raw image. By using geometric transformation, the raw image can be restored and corrected in case of any presence of geometric distortion by projecting each pixel in the raw image onto another space. Wang *et al.* [6] developed a complete AOI system for drilling investigation in the PCB industry. Before comparing or matching standard and testing images, they utilized geometric transformation, including the calculations of scaling, rotating, and translation to correct the direction and angle of the testing image. The results indicated that the proposed AOI system could achieve an accuracy of tiny size defects in drilling inspection. Besides, image denoising and filtering methods are frequently applied for the image containing the ambiguous background, which mainly causes the false positive rate in the recognition process. Using image filtering and denoising can reduce noise and enhance the image for highlighting the important features. Then, the inspection algorithm can perform more effectively by using these highlighted features. Huang *et al.* [20] proposed a defect identification algorithm using a median filter and a clustering approach to identify the clustered defect on the wafer surface. The isolated defective dies on the wafer samples appear as salt-and-pepper noise, they employed a median filter to separate the wafer defect clusters from its

isolated defective dies. Then, the nearest-neighbor method is adopted to identify defective dies associated with various defect clusters. The results showed that the algorithm can effectively detect the clustered defects. Besides, since a cluster is too small, it is treated as noise and removed by the median filter. Therefore, the algorithm does not mistake isolated defective dies for defect clusters. Park and Kweon in [21] proposed a novel filtering method named neighboring difference filter (NDF) to represent ambiguous defects under diverse surface illumination conditions of AMOLED sample images. In the NDF method, the intensity of neighboring regions is compared with a strategically placed gap space to accentuate the differences and remove similarities. Spatial domain methods directly deal with pixel values that have low computation complexity. However, these methods do not provide adequate robustness for defect image enhancement.

The frequency-domain methods used transformations, such as discrete cosine transform (DCT) [22], discrete Fourier transform (DFT) [23], and discrete wavelet transform (DWT) [24], to transform images from the spatial domain to the frequency domain. In the frequency domain, the image is separated into different subbands which contain more detailed components of an image. After processing the image in the frequency domain, the enhanced image is obtained by using invert transform to convert the image from the frequency domain to the spatial domain. Since the solder joint on a PCB image is highly variant with illumination variation, the appearance of the same solder joint can be greatly different. Mar *et al.* [25] overcame this challenge by using DCT to normalize the solder joint images to keep them stable under variant illumination conditions. Meanwhile, Lin and Ho [26] enhanced the pinhole defects on chips and wafers to effectively eliminate the global random texture pattern and emphasize tiny pinhole defects in the enhanced images. The frequency matrix of an image in the DCT domain is decomposed to select the best radius of the sector filter before applying high-pass filtering operations. The results demonstrated that this method is invariant to the orientation of the target chip and obtains high accuracy in pinhole defect detection. To detect visual blemishes of the curved LED lenses, Chiu and Lin [27] firstly divide the input image into non-overlapping blocks of the same sizes and then transform these blocks into the DCT domain to extract the DCT block's representative energy features. Finally, a complex method by combining statistic methods, grey clustering technique, thresholding methods, is used on energy features to distinguish real blemished areas and uniform regions. Wu *et al.* in [19] used DFT firstly to remove background textures and eliminate the influence of uneven illumination of the TF-LCD images. Then, the tiny defect features of TFT-LCD are highlighted before using the support vector machine (SVM) algorithm for defect inspection. Similarly, Tsai and Hung [28] used one-dimensional DFT to transform the TFT-LCD image to the frequency domain and then eliminate the frequency components that represent
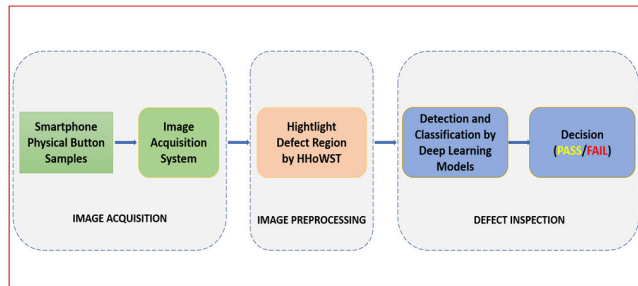
FIGURE 1. The complete AOI system architecture for the smartphone's physical button inspection.

the periodic pattern to remove the patterned background of the TFT-LCD image, whereas local anomalies are preserved for defect inspection on the TFT-LCD panels. Compared to DCT and DFT, DWT is advantageous in providing a simple way for multi-resolution analysis. Lin [29] used Haar wavelet to enhance water-drop defects on the surface of the LED chip. The LED chip image is decomposed by Haar wavelet at the first level to extract wavelet characteristics. Then, they applied the wavelet-based neural network and multivariate statistical techniques to inspect the defect. These methods can clarify the features of the defect on the product image. However, they only focus on different frequency subbands separately, lead to an unsatisfactory inspection recognition rate.

In this paper, we propose a complete AOI system, as shown in Fig. 1, which includes three primary parts for the smartphone's physical button inspection. In the first part, the image acquisition system, we design an effective image acquisition system to take the smartphone's physical button images. By using the designed image acquisition system, we can capture high-resolution images to make defects clearer. However, the defect detection and classification results still do not satisfy our requirements. In the second part, image preprocessing, based on the combination of the wavelet transform and tensor theory, we propose a defect enhancement method called the highlight defect region by using higher-order singular value decomposition of wavelet subband-based tensor (HHoWST) to highlight the defect region on the smartphone's physical button image. In the last part, defect inspection, we adopt state-of-art deep learning models in the defect inspection step to evaluate the improvement of the defect detection rate. The conclusion of the quality of the smartphone's physical button is based on the manufacturer's standards. The experimental results reveal that the detection rate of the AOI system using our proposed method has significant improvement. Especially, the recognition rate of the tiny defects, which are usually challengeable for most AOI systems.

The rest of this paper is organized as follows. Section. II presents the image acquisition system to collect the smartphone button image dataset. Section III details

our proposed method. Section IV presents the experimental results, and Section V presents the conclusions.

II. THE IMAGE ACQUISITION SYSTEM AND SMARTPHONE'S PHYSICAL BUTTON IMAGE DATASET

In this section, we firstly introduce the system architecture of the smartphone's physical button inspection equipment and present appropriate significance regarding the illumination scheme. Next, the smartphone's physical button dataset (DCL-CBI) collected by ourselves is acquired with a total of 500 images of two smartphone's physical button types. Finally, the image properties and the challenges of defect detection are also discussed.

A. THE IMAGE ACQUISITION SYSTEM

This subsection introduces the components of the proposed image acquisition system, such as the light source setting, camera parameter setting, and image capture design. In an image acquisition system, the lighting setup plays a vital role in ensuring the quality of the acquired images. When positive light is incident on an object's surface, some of the light gets reflected based on the refractive indices of the coating material on the object's surface, angle of incidence, and medium of the incident light. The light may be reflected back and forth several times, thus capturing images with variable qualities. Therefore, resolving the issue of how to design a suitable image acquisition system for each type of object is crucial in the AOI system.

In this paper, we design an image acquisition system for two types of smartphone's physical buttons, named as volume up/down button and side button. These physical buttons are small, which the volume up/down button has a size of 110×25 mm^2 and the side button has a size of 165×25 mm^2. Besides, physical buttons also have non-planar and shiny surfaces, which are very sensitive to lighting. To overcome these difficulties, it is necessary to prioritize the influence of the light source and the smartphone's physical buttons to increase the contrast difference between defects and the background. Moreover, the light may be reflected back and forth several times, thus capturing images with variable qualities. In addition, the quality of the acquired image is also affected by unexpected lighting from the surrounding. So, the designed system is placed in a dark room. In addition, the sample is placed on a holder, which is made of black fabric to avoid the unexpected lighting reflection from its surface to the camera. Therefore, the camera only receives the lighting reflected from the surface of the smartphone's physical button. Fig. 2 displays the proposed smartphone's physical buttons image acquisition system. As can be seen in Fig. 2, the camera used in the designed system is BASLER acA4112-30uc with a fixed focal length lens. During the image acquiring process, the lens aperture is constantly set to f/1.4 at the fixed focal length of 35mm. By using the designed image acquisition system, we can capture high-resolution images to make defects clearer.
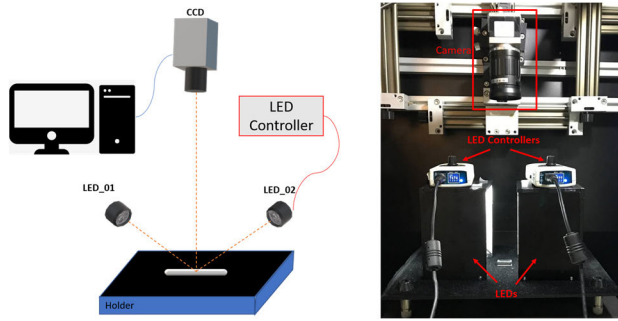
FIGURE 2. The image acquisition camera system for the smartphone's physical button.
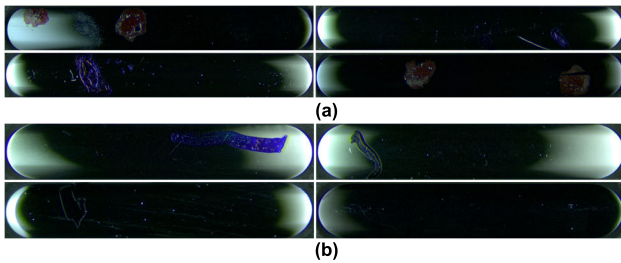

(a)

(b)

FIGURE 3. The smartphone's physical button images in the DCL-CBI dataset. (a) The side button images. (b) The volume up/down button images.

B. SMARTPHON'S PHYSICAL BUTTON IMAGE DATASET

Normally, the smartphone's physical button is small and has non-planar and shiny surfaces. Thus, it is very difficult to detect defects with human eyes. Therefore, image acquisition is a critical step in the defect inspection system. By using our proposed image acquisition system, a total of 500 images from smartphone's physical buttons provided by a manufacturer are captured and constructed a dataset, named as DCL-CBI dataset. Fig. 3 shows four smartphone physical button images in the DCL-CBI dataset. The acquired images are resized to two different sizes, which correspond to two types of the smartphone's physical buttons, to reduce the time processing during the inspection process. In particular, the images are resized to $1,000 \times 180$ pixels for volume up/down button images and $1,600 \times 200$ pixels for side button images with a 24-bit BMP format. These images are not only evaluated for obtaining a high defect detection rate but also reduced in size to decrease the operation time ensuring synchronization with the factory's production line.

After the DCL-CBI dataset is constructed by using the proposed image acquisition system, we manually classify based on the manufacturer's specifications. As shown in Fig. 4, the defects are classified into the following six types—white mist (WM), collapsed side (CS), scratch (SR), pressed mark (PM), small dot (SD), and pen mark (MK). There are huge differences among the sizes of these defect types. The size of the largest defect on the smartphone's physical button surface could be 2×2 mm^2, while the smallest defect is less than 0.1×0.1 mm^2, such as the SD defect type. Moreover, a tiny defect presented in a smartphone's physical button image may only differ slightly from the surrounding region.
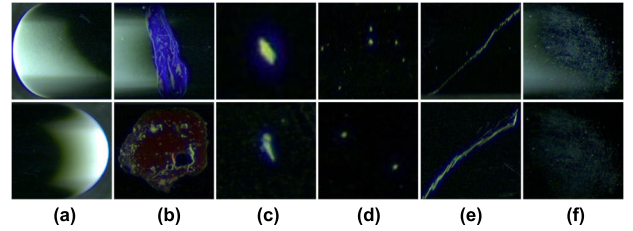

FIGURE 4. The defect types of smartphone's physical button in the DCL-CBI dataset. (a) CS. (b) MK. (c) PM. (d) SD. (e) SR. (f) WM.

Besides, the appearance of the defects is not always uniform on the smartphone's physical button surface. For example, the intensities of some SD defects shown in Fig. 4(d) and SR defects shown in Fig. 4(e) are relatively low. The colors of the MK defect in Fig. 4(b) and WM in Fig. 4(f) are partly suppressed by variant illumination of the smartphone's physical button surface. And therefore, although using the specialized image acquisition system, the collected images still have many challenges to recognize the defect. In this paper, we propose a method called HHoWST to highlight the defect region for improving the recognition rate of the AOI system. The detail is discussed in the next section.

III. HIGHLIGHT DEFECT REGION USING WAVELET SUBBAND-BASED TENSOR

In this section, the proposed method is presented in detail. First, the multiresolution analysis using wavelet transform is briefly discussed. Then, the wavelet subband-based tensor, which is the combination of the wavelet subbands representation and tensor theory, and its higher-order singular value decomposition (HOSVD) [38], are presented. Finally, the proposed method used to eliminate the components, which contain the global illumination information and lead to highlight the defect region, is presented.

A. WAVELET REPRESENTATION

Wavelet transform is a signal transformation, which is widely applied in digital image processing. Specifically, the two-dimensional DWT (2D-DWT) has been studied in the problems of image compression [30], [31], image enhancement [32], and texture representation [33], [34]. By using the scaling and shifting operations, 2D-DWT can effectively extract information from signals and represent a multiscale of signals or functions. The wavelet representation of a 2D-signal $I(x, y)$ at an arbitrary $M-$ level is given by:

$$I(x, y) = \sum_{k,l} c_{k,l}^{M} \Phi_{M,k,l}^{0}(x, y) + \sum_{j=1}^{M} \sum_{\tau \in T} \sum_{k,l} d_{k,l}^{j,\tau} \Psi_{j,k,l}^{\tau}(x, y),$$
(1)

where $j, k \in \mathbb{Z}$ denote the scale and translation parameters, respectively. The approximation coefficients and the detail coefficients are represented as $c_{k,l}^{M} = \left\langle I, \Phi_{M,k,l}^{0} \right\rangle$ and $d_{k,l}^{j,\tau} = \left\langle I, \Psi_{j,k,l}^{\tau} \right\rangle$, where $\langle ., . \rangle$ denotes the inner product in the space $L^2(\mathbb{R}^2)$ and $\tau \in T = \{1, 2, 3\}$. A scaling function

$\Phi_{M,k,l}^{0}(x, y)$ and three wavelet basis functions $\Psi_{j,k,l}^{\tau}(x, y)$ are defined as follows:

$$\Phi_{j,k,l}^{0}(x, y) = \phi_{j,k}(x) \phi_{j,l}(y), \quad (2)$$

$$\Psi_{j,k,l}^{1}(x, y) = \phi_{j,k}(x) \psi_{j,l}(y), \quad (3)$$

$$\Psi_{j,k,l}^{2}(x, y) = \psi_{j,k}(x) \phi_{j,l}(y), \quad (4)$$

$$\Psi_{j,k,l}^{3}(x, y) = \psi_{j,k}(x) \psi_{j,l}(y). \quad (5)$$

The basic functions $\{\phi_{j,k}\}$ and $\{\psi_{j,k}\}$ are also determined by $\phi_{j,k} = 2^{-j/2}\phi(2^{-j}x - k)$ and $\psi_{j,k} = 2^{-j/2}\psi(2^{-j}x - k)$. Therefore, a two-dimensional signal can be approximated by the sum of approximation and detail coefficients at multiresolution scales. In other words, there is a sequence of approximations of a given function $I(x, y)$ at different resolutions [35], [36]. Moreover, these approximations at a resolution 2^j can be defined as the orthogonal projection of $I(x, y)$ on a subspace V_j that satisfies

$$\cdots V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \cdots, \quad (6)$$

with

$$\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}) \quad and \quad \bigcap_{j\in\mathbb{Z}} V_j = \{0\}. \quad (7)$$

Denote W_j as the complement space of V_j. The information about the resolution 2^{j+1} can be decomposed into the approximation in subspace V_j and the details in the subspace W_j. To this end, the orthogonal projection in space V_{j+1} can be written as

$$V_{j+1} = V_0 \oplus [W_0 \oplus W_1 \oplus \cdots \oplus W_j], \quad (8)$$

where V_0 is the subspace at the coarsest scale and the notation $\oplus$ is the direct sum operator that satisfies

$$V_0 \cap W_0 \cap W_1 \cap \cdots \cap W_j = 0. \quad (9)$$

Hence, by applying 2D-DWT, a given image can be decomposed into several multiresolution frequency components called LL_j, HL_j, LH_j, HH_j subbands. The subband denoted LL corresponds to approximation coefficients, whereas HL_j, LH_j, HH_j are the detail coefficients, which contain the changes of images along with vertical, horizontal, and diagonal directions, respectively. Indeed, the decomposition into different frequency ranges (subbands) allows to isolate the frequency components and process each independently as mentioned in Eqs. (8), (9).

As seen in Fig. 5(b), the Daubechies 2D-DWT decomposes a smartphone's physical button image as shown in Fig. 5(a) at the first level into LL, HL, LH, and HH subbands. To demonstrate the influence of the relevant information contained in each subband among such four subbands, the reconstructed images using only one corresponding subband (the remains are set to zero) are also depicted in Figs. 5(c)-5(f). In particular, the LL subband contains a considerable amount of information of the original image since the reconstructed image shown in Fig. 5(c) still retains most of the information related to the illumination
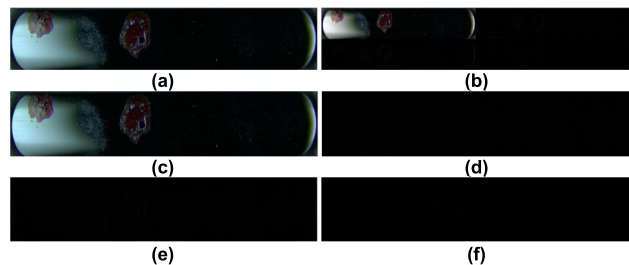

FIGURE 5. Wavelet decomposition and reconstructed images of a smartphone's physical button image from the DCL-CBI dataset. (a) A smartphone physical button image; (b) Four subbands of (a) obtained by using 2D-DWT at the first level; (c) Reconstructed image using only LL subband; (d) Reconstructed image using only HL subband, (e) Reconstructed image using only LH subband; (f) Reconstructed image using only HH subband.

and structure of the button. In contrast, the reconstructed images using HL, LH, and HH subbands, as shown in Figs. 5(d)-5(f), indicate that the middle- and high-frequency subbands refer to the noise and the detail of the image structure corresponding to vertical, horizontal, and diagonal directions, respectively. Indeed, the key features in an image are represented by the larger coefficients that appear almost on the low-frequency subband, while the middle- and high-frequency subband coefficients are quite small and closer to zero.

By using 2D-DWT as a multiresolution image decomposition tool, image information can be spread out and expressed through various frequency subbands separately. This allows to extract the meaning features from the image and process them separately for specific purposes. In this study, we propose a method to construct a tensor called wavelet subbands-based tensor (WST), which is based on the combination of the wavelet representation and tensor theory to express the image information in high-order dimensions.

B. WAVELET SUBBANDS-BASED TENSOR

In this subsection, the WST is introduced. Firstly, the tensor fundamentals are reviewed, and then the WST construction is discussed in detail in the next subsection.

1) TENSOR PRELIMINARIES

Tensor [37] is defined as the multidimensional arrays of numerical values used to represent multidimensional data. Specifically, it is widely applied in representing color images, videos, and hyperspectral images through various data arrangements. To ensure the clarity of the notation, we first introduce some terminology and notations of tensors that are used in this paper [38], [39].

i) An Nth- order tensor denoted as $\mathcal{X}$ is formally defined as an element of the tensor product of N vector spaces, where the order N of a tensor is the number of dimensions and $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$.

ii) *Fibers* are the higher-order analog of matrix rows and columns. A fiber of tensor $\mathcal{X}$ can be determined by keeping an index, whereas fixing others. A third-order tensor $\mathcal{X}$ involves

column, row, and tube fibers denoted as $x_{:ml}$, $x_{n:l}$, and $x_{nm:}$, respectively.

iii) *Slices* are two-dimensional portions of a tensor. A slice is defined by fixing all indices but two. A third-order tensor $\mathcal{X}$ contains slices denoted as $X_{n::}$, $X_{:m:}$, and $X_{::l}$, which are called horizontal, lateral, and frontal slices, respectively.

iv) A tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ has NN ways to transform its data into a matrix called mode-k unfolding of tensor $\mathcal{X}$ or tensor matricization. It is denoted by $X_{(k)}$, where $k \in \{1, 2, \ldots, N\}$. In particular, as defined in [39], the element $(i_1, i_2, \cdots, i_N)$ of tensor $\mathcal{X}$ with size $(I_1 \times I_2 \times \cdots \times I_N)$ maps to matrix element (i_k, j) of $X_{(k)}$, which is the mode-k unfolding of tensor $\mathcal{X}$, where j is computed as:

$$j = 1 + \sum_{\substack{k=1 \\ k \neq n}}^{N} \left[(i_k - 1) \prod_{\substack{m=1 \\ m \neq n}}^{k-1} I_m \right]. \qquad (10)$$

v) The mode-k product is the result of the multiplication of a tensor by a matrix in mode-k. The mode-k product $\mathcal{X} \times_k U$ of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ with a matrix $U \in \mathbb{R}^{J \times I_k}$ is a tensor of size $(I_1 \times \cdots \times I_{k-1} \times J \times I_{k+1} \times \cdots \times I_N)$. Therefore, we have

$$(\mathcal{X} \times_k U)_{i_1 \cdots i_{k-1} j i_{k+1} \cdots i_N} = \sum_{i_k} x_{i_1 i_2 \cdots i_N} u_{j i_k}, \qquad (11)$$

where $x_{i_1 i_2 \cdots i_N}$ and $u_{j i_k}$ are the elementwise of $\mathcal{X}$ and U, respectively.

2) WST CONSTRUCTION

To effectively exploit the relationship between the spatial information, color channels, wavelet subbands of the smartphone's physical button image, we construct a tensor with twelve frontal slices denoting twelve wavelet subbands of three-color channels of the smartphone's physical button image. Given a smartphone's physical button color image I of size $h \times w$, I is split into I^b, I^g, and I^r, which correspond to the three-color channel images in RGB color space. By applying the 2D-DWT at the first level, each color channel image I^c, $c \in \{b, g, r\}$, is decomposed into the corresponding frequency subbands of size $h/2 \times w/2$, which are denoted as LL^c, HL^c, LH^c, and HH^c, as shown in Figs. 6 (b), (c), and (d). At this point, a tensor $\mathcal{X} \in \mathbb{R}^{h/2 \times w/2 \times 12}$ sized $h/2 \times w/2 \times 12$ is constructed by treating each frequency subband of each color channel to be a frontal slide $X_{::l}$, where $l = 1, 2, \ldots, 12$, of tensor $\mathcal{X}$. The frontal slides $X_{::l}$ are arranged in groups based on color channels in order of blue, green, and red channels. In each group, the frequency subbands are sorted in order of low-, middle-, and high-frequency subbands. Finally, the WST three-order tensor $\mathcal{X}$ based on the wavelet frequency subbands of color channels is constructed as demonstrated in Fig. 6(f). The WST as shown in Fig. 6(f) is represented by a third-order tensor which is defined by two indices (vertical and horizontal) for spatial variables and the remaining index for the wavelet subbands

at the first level scale of color channels. Hence, the spatial information of the given image is represented along vertical and horizontal axes, whereas the correlations between color channels and frequency subbands are represented by the wavelet subbands of the color channels axis. The construction of WST is summarized in Algorithm 1.

Algorithm 1 WST

Input: Color image $I \in \mathbb{R}^{h \times w}$ size of $h \times w$.
Output: Tensor $\mathcal{X} \in \mathbb{R}^{h/2 \times w/2 \times 12}$.

1. $I^{(c)}$, $c \in (b, g, r) \leftarrow split\ (I)$;
2. $\{LL^c, HL^c, LH^c, HH^c\} \leftarrow$ 2D-DWT (I^c);
3. Construct tensor $\mathcal{X} \in \mathbb{R}^{h/2 \times w/2 \times 12}$ by
   $X_{::l} \leftarrow \{LL^c, HL^c, LH^c, HH^c\}$, $l \in \{1, 2, \cdots, 12\}$;
4. Return $\mathcal{X}$.

The WST can be decomposed by using higher-order tensor decomposition methods, such as PARAFAC [40] or TUCKER [41]. The PARACFAC method is commonly applied for estimation and signal modeling purposes, as the data can often be decomposed into individual components uniquely. Tucker decomposition (or higher-order singular value decomposition-HOSVD) is a well-known higher-order tensor decomposition method and more suited for data compression, tensor approximation, and image enhancement application. In this paper, we firstly use HOSVD to decompose the WST into more detailed components. Next, we concentrate on analyzing them to propose a method to highlight the defect feature of the smartphone's button image in the DCL-CBI dataset to improve the recognition rate in an AOI system.

C. HIGHER-ORDER SINGULAR VALUE DECOMPOSITION OF WST (HoWST)

The tensor is expressed as a linear combination of the outer products in the different modes. In particular, HOSVD expresses an N-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ as

$$\mathcal{X} = \mathcal{S} \times_1 U^{(1)} \times_2 U^{(2)} \times_3 \cdots \times_N U^{(N)}, \qquad (12)$$

or, elementwise, as

$$x_{i_1 i_2 \cdots i_N} = \sum_{j_1=1}^{I_1} \sum_{j_2=1}^{I_2} \cdots \sum_{j_N=1}^{I_N} s_{j_1 j_2 \cdots j_N} u_{i_1 j_1}^{(1)} u_{i_2 j_2}^{(2)} \cdots u_{i_N j_N}^{(N)}, \qquad (13)$$

where $\mathcal{S} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is called core tensor and the matrices $U^{(k)} \in \mathbb{R}^{I_k \times I_k}$, $k = \{1, 2, \cdots, N\}$ are matrices containing the basis of the left singular vectors of the mode-k unfolding matrices $X_{(k)}$ of tensor $\mathcal{X}$. The matrices $U^{(k)}$, named as inverse factors, can be found directly through the implementation of singular value decomposition of mode-k unfolding $X_{(k)}$ as follows:

$$X_{(k)} = U^{(k)} \sum^{(k)} V^{(k)T}, \qquad (14)$$

where $\sum^{(k)}$ and $V^{(k)}$ are the singular value matrices and the right singular vector matrices of $X_{(k)}$, respectively. Besides,
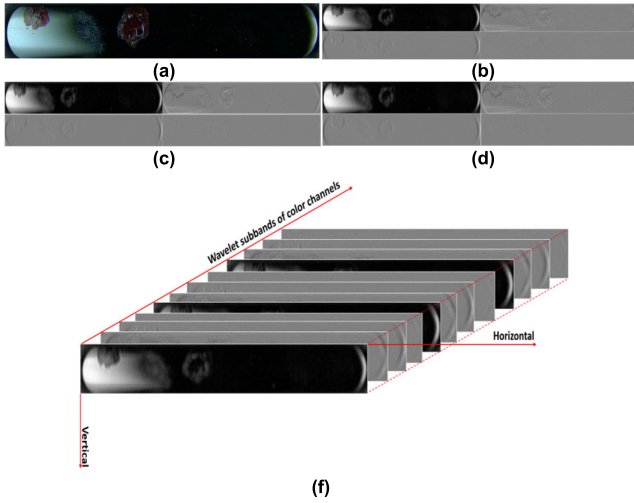
FIGURE 6. (a) An image in the DCL-CBI dataset; (b) Four wavelet subbands of the blue channel; (c) Four wavelet subbands of the green channel; (d) Four wavelet subbands of the red channel; (f) The three-order tensor of (a).

the core tensor $\mathcal{S}$ can be determined by

$$\mathcal{S} = \mathcal{X} \times_1 U^{(1)T} \times_2 U^{(2)T} \times_3 \cdots \times_N U^{(N)T}. \quad (15)$$

Therefore, the mode-k unfolding of $\mathcal{S}$ can be obtained by

$$S_{(k)} = \sum\nolimits^{(k)} V^{(k)} \left(U^{(N)} \otimes \cdots \otimes U^{(k+1)} \right.$$
$$\left. \otimes U^{(k-1)} \otimes \cdots \otimes U^{(1)} \right), \quad (16)$$

where $\otimes$ denotes the Kronecker product of two matrices. Finally, the mode-k unfolding $X_{(k)}$ matrix of the given tensor $\mathcal{X}$ can be reconstructed by

$$X_{(k)} = U^{(k)} S_{(k)} \left(U^{(N)} \otimes \cdots \otimes U^{(k+1)} U^{(k-1)} \otimes \cdots \otimes U^{(1)} \right)^T. \quad (17)$$

As discussed in Sub-section III.B2, given a smartphone's physical button color image I of size $h \times w$, the WST of I is represented by a third-order tensor $\mathcal{X} \in \mathbb{R}^{h/2 \times w/2 \times 12}$, as shown in Fig. 6(f). WST contains both spatial information, and correlations between color channels and frequency subbands of a color smartphone's physical button image. WST $\mathcal{X}$ includes 12 frontal slices, denoted by $X_{::l}$, $l = 1, 2, \ldots, 12$. By using HOSVD, $\mathcal{X}$ is decomposed into 4 components: three inverse factors $U^{(1)}$, $U^{(2)}$, and $U^{(3)}$ and core tensor $\mathcal{S}$. The inverse factors and core tensor $\mathcal{S}$ are presented in Fig. 7. Accordingly, inverse factors $U^{(1)}$, $U^{(2)}$, and $U^{(3)}$ in Figs. 7(a), (b), (c) are matrices of size $h/2 \times h/2$, $w/2 \times w/2$, and 12×12, respectively, whereas core tensor $\mathcal{S}$ is a three-order tensor of size $h/2 \times w/2 \times 12$. Therefore, $\mathcal{S}$ also contains 12 frontal slices $S_{::l}$, $l \in \{1, 2, \cdots, 12\}$, and can be unfolded at mode-k into $S_{(k)}$, $k \in \{1, 2, 3\}$ matrices. The detailed implementation of applying HOSVD to decompose WST (HoWST) is summarized in Algorithm 2.

To determine the specific information carried by HoWST components, as shown on the right side of Eq. 12, we conducted several experiments on the image in the DCL-CBI

Algorithm 2 HoWST

Input: WST $\mathcal{X} \in \mathbb{R}^{h/2 \times w/2 \times 12}$
Output: Inverse factors $U^{(1)}, U^{(2)}, U^{(3)}$ and core tensor $\mathcal{S}$

1. $X_1, X_2, X_3 \leftarrow unfolding\,(\mathcal{X})$;
2. $get(U^{(k)}, \sum^{(k)}, V^{(k)})$ via Eq. (14);
3. $get(S_{(k)})$ via Eq. (16);
4. $\mathcal{S} \leftarrow$ Reshape $(S_{(k)})$
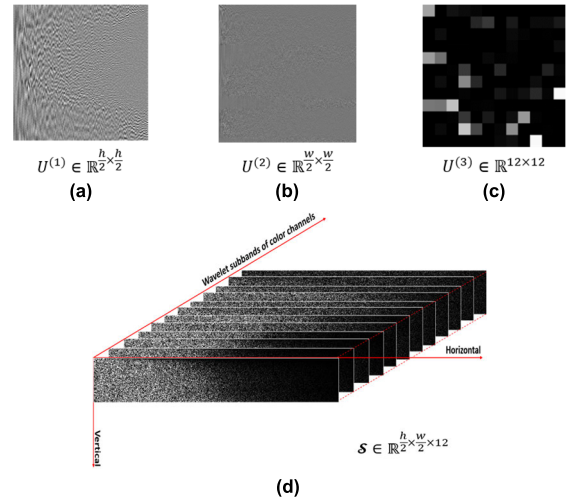4.   Return $U^{(1)}, U^{(2)}, U^{(3)}, \mathcal{S}$.

FIGURE 7. The HOSVD components. (a) Inverse factor $U^{(1)}$; (b) Inverse factor $U^{(2)}$; (c) Inverse factor $U^{(3)}$; (d) Core tensor $\mathcal{S}$.

dataset. The idea is to remove the effect of the investigated component in the result tensor reconstructed by using Eq. (17). As mentioned in Eq. (12) and Eq. (13), the given tensor can be expressed by the linear combination of the outer products of the basis vectors. Therefore, the investigation of the HoWST components can be conducted by alternately replacing the core tensor $\mathcal{S}$ with the all-one tensor and inverse factors $U^{(i)}$, $i = 1, 2, 3$ by the all-one matrices, whereas keeping the others. Thus, the reconstructed tensor is the combination of all HoWST components except the investigated component. Recalled that an all-one matrix is a matrix in which all values are 1. Similarly, the all-one tensor is a tensor in which all frontal slices are all-one matrix. By replacing one-by-one $U^{(1)}$, $U^{(2)}$, $U^{(3)}$, and $\mathcal{S}$ with all-one matrix and all-one tensor respectively, the reconstructed tensors denoted as $\mathcal{X}_{U^{(1)}}$, $\mathcal{X}_{U^{(2)}}$, $\mathcal{X}_{U^{(3)}}$, and $\mathcal{X}_S$ can be expressed elementwise respectively as

$$x_{(u_1)i_1 i_2 i_3} = \sum_{j_1=1}^{I_1} \sum_{j_2=1}^{I_2} \sum_{j_3=1}^{I_3} s_{j_1 j_2 j_3} u_{i_2 j_2}^{(2)} u_{i_3 j_3}^{(3)}, \quad (18)$$

$$x_{(u_2)i_1 i_2 i_3} = \sum_{j_1=1}^{I_1} \sum_{j_2=1}^{I_2} \sum_{j_3=1}^{I_3} s_{j_1 j_2 j_3} u_{i_1 j_1}^{(1)} u_{i_3 j_3}^{(3)}, \quad (19)$$

$$x_{(u_3)i_1 i_2 i_3} = \sum_{j_1=1}^{I_1} \sum_{j_2=1}^{I_2} \sum_{j_3=1}^{I_3} s_{j_1 j_2 j_3} u_{i_1 j_1}^{(1)} u_{i_2 j_2}^{(2)}, \quad (20)$$
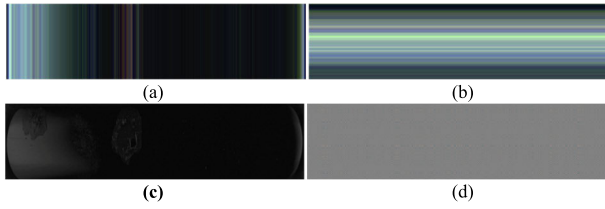
FIGURE 8. The reconstructed image of Fig. 6(a) by replacing WST's HOSVD components with all-one matrix or all-one tensor, respectively. (a) Replacing $U^{(1)}$ with all-one matrix; (b) Replacing $U^{(2)}$ with all-one matrix; (c) Replacing $U^{(3)}$ with all-one matrix; (d) Replacing $\mathcal{S}$ with all-one tensor.

$$x_{(S)i_1 i_2 i_3} = \sum_{j_1=1}^{I_1} \sum_{j_2=1}^{I_2} \sum_{j_3=1}^{I_3} u_{i_1 j_1}^{(1)} u_{i_2 j_2}^{(2)} u_{i_3 j_3}^{(3)}. \tag{21}$$

Fig. 8 demonstrates the result images reconstructed from $\mathcal{X}_{U^{(1)}}$, $\mathcal{X}_{U^{(2)}}$, $\mathcal{X}_{U^{(3)}}$, and $\mathcal{X}_S$ of the given image in Fig. 6(a). As shown in Figs. 8(a) and 8(b), the images reconstructed without the effect of $U^{(1)}$ or $U^{(2)}$ by replacing $U^{(1)}$ or $U^{(2)}$ with all-one matrix, still keeps the color information along with vertical and horizontal directions. Nevertheless, the structure of the smartphone's physical button image is destroyed completely. Hence, $U^{(1)}$ and $U^{(2)}$ contain most structure information of the given image. Similarly, as seen in Fig. 8(c), the color of the reconstructed image without $U^{(3)}$ is removed completely, and the image is converted to a gray image. Furthermore, a small amount of structure of the image is also discarded. Thus, $U^{(3)}$ does contain not only the color information but also a small amount of information related to the structure of the image. In Fig. 8(d), the structure and color information of the image is mostly removed. Therefore, the core tensor $\mathcal{S}$ contains a part of information related to the color and structure of the image. Thus, we analyze in more detail the effectiveness of global illumination affected by core tensor $\mathcal{S}$ and inverse factor $U^{(3)}$.

Next, the core tensor $\mathcal{S}$ is further analyzed to investigate the detailed information carried by its components. In particular, to understand the specific information contained in each frontal slice $S_{::l}$, $l = \{1, 2, \cdots, 12\}$ of $\mathcal{S}$, a new core tensor denoted as $\mathcal{S}_f$, $f = \{1, 2, \cdots, 12\}$ is constructed by keeping the coefficients of the fth frontal slice while the remaining slices are set as zero. Therefore, $\mathcal{S}_f$ contains frontal slices $S'_{::f}$, where

$$S'_{::f} = \begin{cases} S_{::l}, & \text{if } f = l \\ 0, & \text{otherwise.} \end{cases} \tag{22}$$

Without loss of generality, a tensor denoted as $\mathcal{X}_{S_f}$ is reconstructed from the core tensor $\mathcal{S}_f$ and the inverse factors $U^{(1)}$, $U^{(2)}$, and $U^{(3)}$, and the image corresponding to tensor $\mathcal{X}_{S_f}$ is also reconstructed by 2D-iDWT and denoted as I_{S_f}, $f = 1, 2, \cdots, 12$. Fig. 9 shows the reconstructed smartphone's physical image from the given images as shown in Fig. 6(a). Fig. 9(a) presents the reconstructed image by using only the first frontal slice of the core tensor $\mathcal{S}$ and the remaining frontal slices are set to zero matrices. As seen, the reconstructed image still keeps global illumination of the
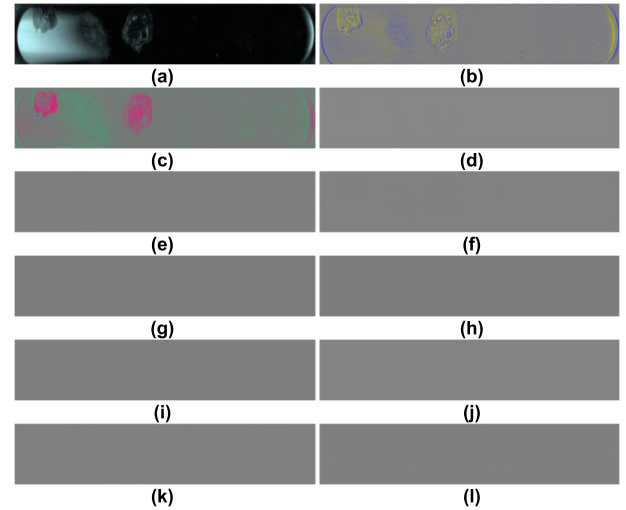

FIGURE 9. The reconstructed images of Fig. 6(a) by keeping only a frontal slice of core tensor and setting the remaining frontal slice to 0 matrices. From (a) to (l) show the reconstructed images corresponding to a frontal slice from 1$^{\text{st}}$ to 12$^{\text{th}}$ frontal slice.

original smartphone's physical button image. Figs. 9(b)-9(c) show the reconstructed images based on the second or third frontal slices of the core tensor $\mathcal{S}$. As seen in Figs. 9(b)-9(c), the global illumination has been removed, the defect regions on the smartphone's physical button image have been highlighted. Figs. 9(d)-9(l) show the reconstructed images by using only one frontal slice, from 4$^{\text{th}}$ to 12$^{\text{th}}$ frontal slices in the core tensor $\mathcal{S}$. As seen, the last nine frontal slices carry less lighting and color information, while mainly contain information regarding the structure detail and noise of the image along with various directions.

We continue analysis more detail on the elements of the first frontal slice matrix. The element values of the first frontal slice matrix are large, so we do experiments by keeping only the maximum value of the first frontal slice matrix while the remaining values are set to 0. The reconstructed image is shown in Fig. 10(b). As seen, the maximum value of the first frontal slice matrix still maintains the global illumination on the surface of the smartphone's physical button image. Fig. 10(c) shows the reconstructed image of Fig. 10(a) by setting the maximum value of the first frontal slice matrix to 1 and keeping the remaining values. As seen, the global illumination on the surface of a smartphone's physical button image has been estimated. Therefore, the defect region can be highlighted by eliminating the first frontal slice and keeping the remaining frontal slices to remove the influence of global illumination.

The relevant information carried by the subcomponents of $U^{(3)}$ is further analyzed in detail by fixing the core tensor $\mathcal{S}$, inverse factors $U^{(1)}$ and $U^{(2)}$ but changing $U^{(3)}$. To understand the specific information contained in each row of the inverse factor matrix $U^{(3)}$, we fix the values of a row and set the remaining rows to 0. The modified $U^{(3)}$ matrix based on values of the jth row denoted as $U_f^{(3)}$,
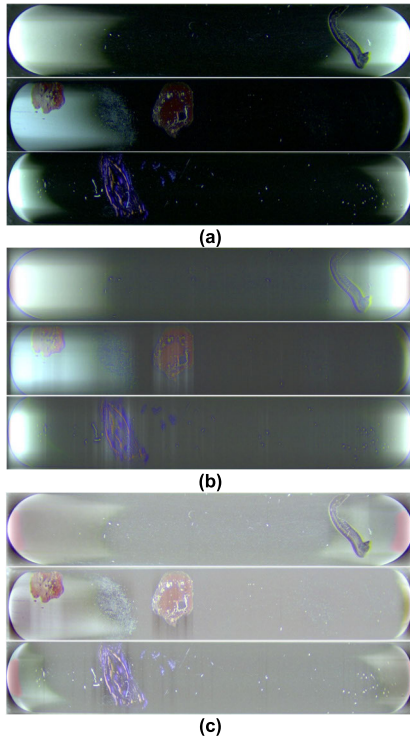
FIGURE 10. Analyzing the effect of the maximum element of the first frontal slice matrix. (a) Three original images; (b) The reconstructed images by keeping only the maximum value of the first frontal slice matrix and the remaining values are set to 0; (c) The reconstructed image by setting the maximum value of the first frontal slice matrix to 1 and keeping the remaining values.
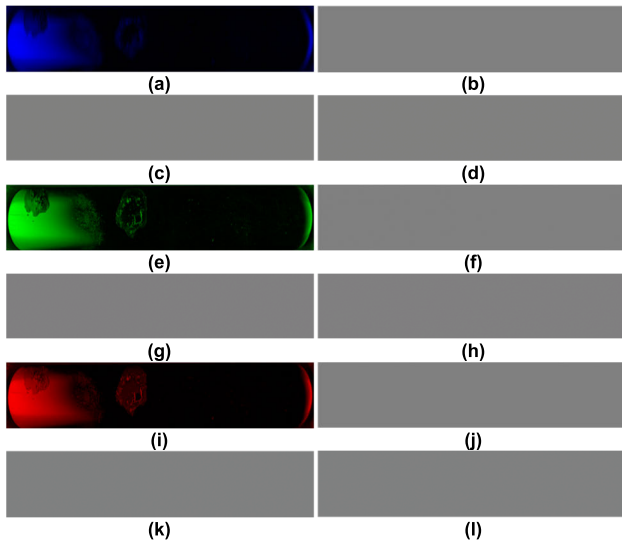

FIGURE 11. The reconstructed images of Fig. 6(a) by keeping a row of $U^{(3)}$ and setting the remaining rows to 0. From (a) to (l) show the reconstructed images corresponding to a row from 1st to 12th row.

$f = 1, 2, \cdots, 12$, is determined as

$$U^{(3)}_{f=1,2,\cdots,12}(i,j) = \begin{cases} U^{(3)}(i,j), & \text{if } i = f \\ 0, & \text{otherwise.} \end{cases} \quad (23)$$

The reconstructed images based on the $U^{(3)}_f$, the core tensor $\mathcal{S}$, the remaining inverse factors $U^{(1)}$ and $U^{(2)}$, denoted as

$I_{U^{(3)}_f}$, are shown in Fig. 11. In particular, Figs. 11(a)-11(d) present the reconstructed images $I_{U^{(3)}_f}, f = 1, 2, 3, 4$, based on the 1st, 2nd, 3rd, and 4th rows, which mostly retain the information related to the blue color channel. Fig. 11(a) shows that the 1st row of $U^{(3)}$ contains the global illumination of the blue color channel, whereas images $I_{U^{(3)}_f}, f = 2, 3, 4$, as shown in Figs. 11(b)-11(d), retain the detailed information along with horizontal, vertical, and diagonal directions of the blue channel of the given image, respectively. Thus, the 1st-4th rows can reconstruct the blue channel image information corresponding to the information that is reconstructed by the wavelet subbands LL^b, HL^b, LH^b, and HH^b of the blue channel image. Likewise, Figs. 11(e)-11(h) present the resulting image based on the 5th-8th rows of $U^{(3)}$, which are referred to the information of the wavelet subbands LL^g, HL^g, LH^g, and HH^g of the green channel image. Figs. 11(i)-11(l) present the reconstructed image based on the 9th-12th rows of $U^{(3)}$ which are referred to the information of the wavelet subbands LL^r, HL^r, LH^r, and HH^r of the red color channel image of the given image. Figs. 11(e) and 11(i) show that the 5th and 9th rows of $U^{(3)}$ contain the global illumination of the green and red color channels, respectively. So, we focus on the 1st, 5th, and 9th rows of $U^{(3)}$ to eliminate global illumination of the original smartphone's physical button image to highlight the defect region.

We also continue analysis more detail on the effect for the 1st, 5th, and 9th rows of $U^{(3)}$ by keeping a value of these rows and setting the remaining values of $U^{(3)}$ to 0. Fig. 12 presents the reconstructed images of Fig. 6(a) by keeping one value of a row (one of 1st, 5th, and 9th rows) and setting the remaining values of $U^{(3)}$ to 0. Fig. 12(a) shows twelve reconstructed images by keeping one element of the 1st row and the remaining values of $U^{(3)}$ set to 0. Fig. 12(b) shows twelve reconstructed images by keeping one element of the 5th row and the remaining values of $U^{(3)}$ set to 0. Similarly, Fig. 12(c) shows twelve reconstructed images by keeping one element of the 9th row and the remaining values of $U^{(3)}$ set to 0. As seen, the first values of these rows, such as $U^{(3)}(1, 1)$, $U^{(3)}(5, 1)$, and $U^{(3)}(9, 1)$, contain the color channel information on the entire surface of the smartphone's physical button, whereas the other values keep the illumination on the defect region. Therefore, we estimate the first element of the 1st, 5th, and 9th rows of $U^{(3)}$ to remove the global illumination of the smartphone's physical button image. The defect region is highlighted by estimating the values on the first frontal slice of core tensor $\mathcal{S}$ and the values of the 1st, 5th, and 9th rows of $U^{(3)}$. The detail is discussed in the next subsection.

D. HIGHLIGHT DEFECT REGION

The lighting plays an important role to highlight the defects during the smartphone's physical button image acquiring process. However, different types of defects will have different reflection lighting, lead to the difficulty to highlight all types of defects on the smartphone's physical button
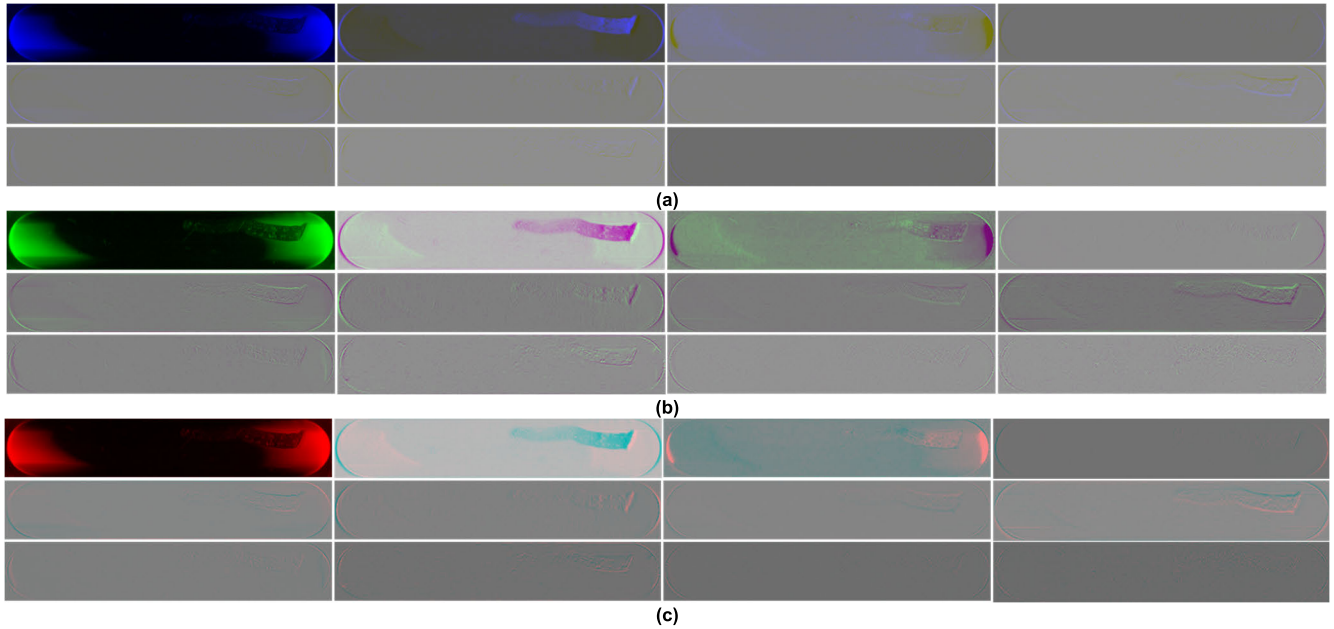
FIGURE 12. The reconstructed images by keeping one element and setting the remaining elements of $U^{(3)}$ to 0. (a) Keeping one element on 1st row; (b) Keeping one element on 5th ro; (c) Keeping one element on 9th row.

image. By using a specialized camera system as mentioned in Section. II, the defect which is very sensitive to the lighting, such as CS type, can be captured. However, the defect appearing on the rest of the sample is harder to determine since the lighting is less reflected from this surface area to the camera.

Without loss of generality, from a given smartphone's physical button color image I of size $h \times w$, a WST $\mathcal{X} \in \mathbb{R}^{h/2 \times w/2 \times 12}$ is first constructed by applying Algorithm 1. Then, $\mathcal{X} = \{X_{::l}, l = 1, 2, \cdots, 12\}$, where $X_{::l}$ are frontal slices of $\mathcal{X}$, is decomposed by HoWST to obtain inverse factors $U^{(1)}$, $U^{(2)}$, $U^{(3)}$, and core tensor $\mathcal{S} = \{S_{::l}; l = 1, 2, \cdots, 12\}$ by applying Algorithm 2. Our purpose is to remove the global illumination of the smartphone's physical button image and highlight the defect region automatically. In this study, we focus on values on the first frontal slice of the core tensor $\mathcal{S}$ and the values of the 1st, 5th, and 9th rows of $U^{(3)}$. As discussed in subsection III.C, the core tensor $\mathcal{S}$ contains a part of information related to the color and structure of the image. Besides, the maximum value of the first frontal slice matrix contains the global illumination on the surface of the smartphone's physical button image. When setting the maximum value of the first frontal slice matrix to 1 and keeping the remaining values, the global illumination on the surface of a smartphone's physical button image has been estimated. However, some structure information of the image is also removed, as shown in Figs. 10(b)-10(c). Therefore, to maintain the structure and remove the global illumination of the smartphone's physical button image, we divide values of the first frontal slice of the core tensor $\mathcal{S}$ by its maximum value. This division remains the structure of the image because this transform is linear. Thus, the maximum of the first frontal slice of the core tensor $\mathcal{S}$ is

calculated as

$$mS_{(1)} = max\left(s_{ij1}\right); \quad i = 1, 2, \cdots, \frac{h}{2}; j = 1, 2, \cdots, \frac{w}{2}. \tag{24}$$

The coefficients of the first frontal slice $S_{::1}$ of the core tensor $\mathcal{S}$, denoted as s_{ij1}, are updated by dividing these coefficients by $mS_{(1)}$ as follows,

$$s_{ij1}^{(u)} = \frac{s_{ij1}}{mS_{(1)}}; \quad i = 1, 2, \cdots, \frac{h}{2}; j = 1, 2, \cdots, \frac{w}{2}, \tag{25}$$

Then, the updated core tensor $\mathcal{S}^{(u)}$ based on the updated first frontal slice, denoted as $S_{::1}^{(u)}$, is obtained as

$$\mathcal{S}^{(u)} = \left\{ S_{::1}^{(u)}, S_{::2}, S_{::3}, \cdots, S_{::12} \right\}. \tag{26}$$

For the effect of $U^{(3)}$, the first values of the 1st, 5th, and 9th rows of $U^{(3)}$ contain the most color information of three color channels of the image and the remaining values contain the most illumination of the defect region. $U^{(3)}$ represents the correlations among color channels by arranging the wavelet subbands of the color channels axis. Thus, changing values of $U^{(3)}$ can convert the illumination of the image. So, we estimate the first values of the 1st, 5th, and 9th rows of $U^{(3)}$ by subtracting values of the 1st, 5th, and 9th rows of $U^{(3)}$, denoted as $U^{(3)}(1, 1)$, $U^{(3)}(5, 1)$, and $U^{(3)}(9, 1)$, from the mean value of these three values. The computing of mean value also maintains the relationship of three color channels of a smartphone's physical button image. The mean value of $U^{(3)}(1, 1)$, $U^{(3)}(5, 1)$, and $U^{(3)}(9, 1)$ is calculated as

$$\mu = \frac{1}{3}\left(U^{(3)}(1, 1) + U^{(3)}(5, 1) + U^{(3)}(9, 1) \right), \tag{27}$$

and we obtained the updated inverse factor $U_{(u)}^{(3)}$ as

$$U_{(u)}^{(3)}(i,j)_{j=1,2,\cdots,12} = \begin{cases} U^{(3)}(i,j) - \mu, & \text{if } i = 1, 5, 9 \\ 0, & \text{otherwise.} \end{cases} \quad (28)$$

To compute the updated color smartphone's button image based on the updated core tensor $\mathcal{S}^{(u)}$ and inverse factor $U_{(u)}^{(3)}$, we first transform the core tensor $\mathcal{S}^{(u)}$ into mode-k unfolding $S_{(k)}^{(u)}$, $k = 1, 2, 3$. Then, the mode-k unfolding $X_{(k)}^{(u)}$ of the updated tensor, denoted as $\mathcal{X}^{(u)}$, is obtained by using one of the following three equations:

$$X_{(1)}^{(u)} = U^{(1)} S_{(1)}^{(u)} \left(U_{(u)}^{(3)} \otimes U^{(2)} \right)^T, \quad (29)$$

$$X_{(2)}^{(u)} = U^{(2)} S_{(2)}^{(u)} \left(U_{(u)}^{(3)} \otimes U^{(1)} \right)^T, \quad (30)$$

$$X_{(3)}^{(u)} = U_{(u)}^{(3)} S_{(3)}^{(u)} \left(U^{(2)} \otimes U^{(1)} \right)^T, \quad (31)$$

where $\otimes$ denotes the Kronecker product. Updated tensor $\mathcal{X}^{(u)} = \left\{ X_{::l}^{(u)}, l = 1, 2, \cdots, 12 \right\}$, where $X_{::l}^{(u)}$ are frontal slices of $\mathcal{X}^{(u)}$, is constructed by reshaping one of mode-k unfolding $X_{(k)}^{(u)}$, $k = 1, 2, 3$. Then, the updated wavelet subbands are obtained by separating the frontal slices of tensor $\mathcal{X}^{(u)}$ as follows:

$$\left\{ LL_u^c, HL_u^c, LH_u^c, HH_u^c \right\}_{c \in \{b,g,r\}} \leftarrow X_{::l,l=1,2,\cdots,12}^{(u)}. \quad (32)$$

The updated color channel images denoted as I_u^c, is reconstructed by applying 2D-i DWT of updated wavelet subbands $\left\{ LL_u^c, HL_u^c, LH_u^c, HH_u^c \right\}_{c \in \{b,g,r\}}$. Finally, merging three updated color channel images $I_u^c, c \in \{b, g, r\}$ to obtain the updated image denoted as I_H, which removes the global illumination on the surface of the smartphone's physical button and highlights the defect region. Fig. 13 illustrates the results of using the proposed method in the DCL-CBI dataset. As seen, the illumination intensities of an SR defect and several SD defects appearing on the surface of the button as shown in Fig. 13(a) are relatively low. It leads to the low contrast of these defects compared to the background of the button. In contrast, the appearance of these defects is clearer, and the contrast is improved after applying the proposed method, as depicted in Fig. 13(b). Similarly, the color of the defect on the button surface in Fig. 13(c) is completely suppressed by the unexpected illumination effect. The color information of this defect is recovered sufficiently by applying the proposed method, as shown in Fig. 13(d). Then, the defect can be easily classified as the MK defect, whereas its appearance in the raw image is not clear, and it leads to a false decision in the classification process. The efficiency of the proposed method is also demonstrated in Figs. 13(f) and 13(h), which are the result images of the proposed method for raw images in Figs. 13(e) and 13(g), respectively. As seen, the background of the button surface is brightened up after applying the proposed method, while the color of the MK defects existing on the button surface is recovered. Moreover, the color of these defects is more uniform between the dark and bright areas. Therefore,

by applying the proposed method, the global illumination has been removed, the defect regions on the smartphone's physical button image have been highlighted. In our proposed method, the Daubechies wavelet D4 transform is used for both the WST tensor constructing and reconstructing. The highlight defect region on the smartphone's physical button image by applying HOSVD of WST (HHoWST) is briefly expressed in Algorithm 3.

Algorithm 3 HHoWST

Input: Smartphone's physical button image $I \in \mathbb{R}^{h \times w}$
Output: Highlighted defect region image $I_H \in \mathbb{R}^{h \times w}$
1. $\mathcal{X} \in \mathbb{R}^{h/2 \times w/2 \times 12} \leftarrow$ Algorithm 1 (I);
2. $\left\{ U^{(1)}, U^{(2)}, U^{(3)}, \mathcal{S} \right\} \leftarrow$ Algorithm 2 ($\mathcal{X}$);
3. $mS_{(1)} \leftarrow$ Calculate the maximum coefficient of the first frontal slice $S_{::1}$ of $\mathcal{S}$ via Eq. (24);
4. $S_{::1}^{(u)} \leftarrow$ Update first frontal slice $S_{::1}$ via Eq. (25);
5. $\mathcal{S}^{(u)} \leftarrow$ Update core tensor $\mathcal{S}$ via Eq. (26);
6. $\mu \leftarrow$ Calculate mean value of $U^{(3)}(1, 1)$, $U^{(3)}(5, 1)$, and $U^{(3)}(9, 1)$ via Eq. (27);
7. $U_{(u)}^{(3)} \leftarrow$ Update inverse factor $U^{(3)}$ via Eq. (28);
8. $S_{(k)}^{(u)} \leftarrow$ Transform $\mathcal{S}^{(u)}$ into mode-k unfolding;
9. $X_{(k)}^{(u)} \leftarrow$ Calculate mode-k unfolding of the updated tensor via Eqs. (29) - (31);
10. $\mathcal{X}^{(u)} = \left\{ X_{::l,l=1,2,\cdots,12}^{(u)} \right\} \leftarrow$ Reshape $\left(X_{(k)}^{(u)} \right)$ to get frontal slices of the updated tensor;
11. $\left\{ LL_u^c, HL_u^c, LH_u^c, HH_u^c \right\}_{c \in \{b,g,r\}} \leftarrow$ Update wavelet subbands via Eq. (32);
12. $I_H^c \leftarrow$ 2D-iDWT $\left(\left\{ LL_u^c, HL_u^c, LH_u^c, HH_u^c \right\}_{c \in \{b,g,r\}} \right)$;
13. $I_H \leftarrow$ Merge $\left(I_H^b, I_H^g, I_H^r \right)$;
14. Return HHoWST image.

We used HHoWST to highlight the defect region on the smartphone's physical button surface image automatically. The effectiveness of the proposed method is discussed in the next section.

IV. EXPERIMENTAL RESULT AND DISCUSSION

In this section, to illustrate the effectiveness of the HHoWST method on defect detection tasks, the evaluation of the experiments adopts color smartphone's physical button images obtained by our lab namely the DCL-CBI dataset. For the defect detection task, the state-of-the-art object detection deep learning models SSD [42], Faster R-CNN [43], and YOLOv5 [44], are used to evaluate the accuracy of the defect detection task. Besides, to demonstrate the efficiency of the proposed HHoWST method, we also did more experiments to compare with a method named adaptive singular value decomposition in the wavelet domain (ASVDW) by Wang *et al.* [32]. ASVDW is an effective image enhancement method that is adaptively computed based on the distributions of the brightness pixels in the three color channels of the colored image and the correlations among their wavelet subbands. Since the illumination of many defects existing on the
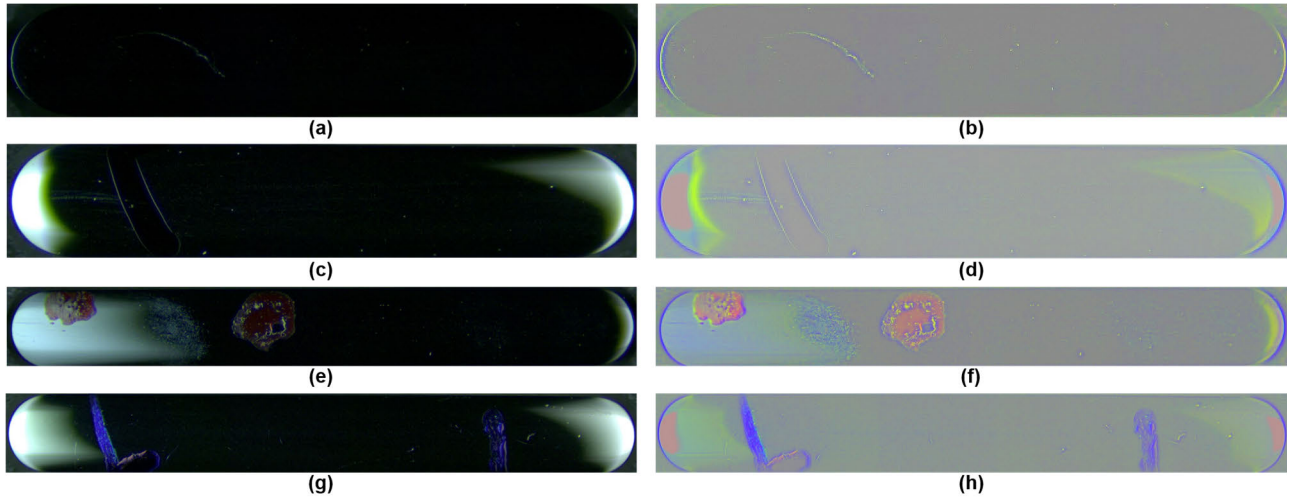
FIGURE 13. Result images are obtained by using the proposed method. The first column shows the original images in the DCL-CBI dataset. The second column shows the results of our method.
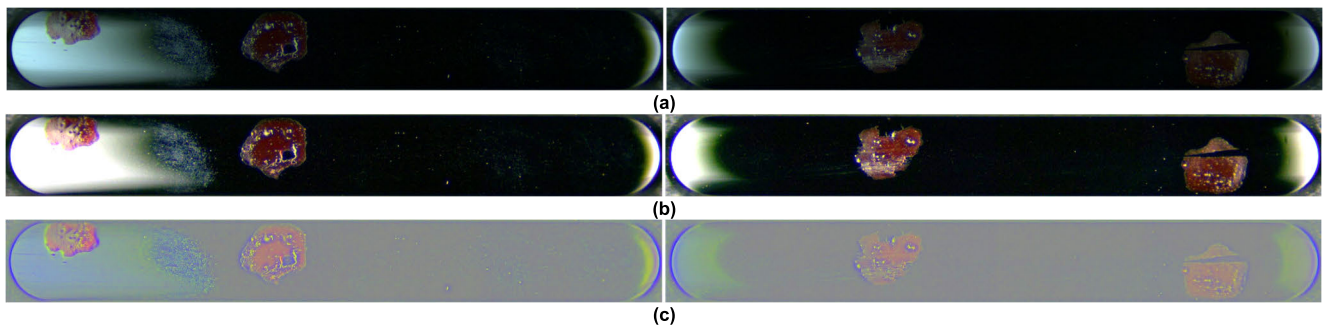

FIGURE 14. The result images by using the proposed HHoWST method and ASVDW method. (a) Two original images from the DCL-CBI dataset. (b) ASVDW image; (c) HHoWST images.

surface of the smartphone's physical button is relatively low, it can be enhanced by using ASVDW. Therefore, the proposed HHoWST method and ASVDW method are both adopted on the color smartphone's physical button images to get the highlight defect region images and the illumination-compensated images, respectively. Then, deep learning models are used to detect the defects on original, ASVDW, and HHoWST images for comparison. The experiments in this paper were implemented on a workstation with an Intel Core i7-9700, RAM 32GB, a single GPU NVIDIA GTX 1660 SUPER on Windows 10.

In our dataset, a total of 500 images of the smartphone's physical buttons provided by a manufacturer are captured by our designed image acquisition system, as discussed in Section II. A. There are two types of smartphone's physical buttons, one is the volume up/down button with a size of 110×25 mm^2 and the other is the side button with a size of 165×25 mm^2. The volume up/down button type has 327 images sized $1,000 \times 180$ pixels and the side button type has 173 images sized $1,600 \times 200$ pixels. Based on the manufacturer specification, we manually classify the defects into the following six classes WM, CS, SR, PM, SD, and MK. A total of 13,472 defects are manually labeled, as shown in Table 1.

TABLE 1. Number of defects in the DCL-CBI dataset.

	WM	CS	SR	PM	SD	MK
No. of defects	36	320	1,910	1,395	9,505	306

There are several challenges to this dataset. First, the size defects have a large difference. The largest defect has a size of 2×2 mm^2, while the smallest defect has a tiny size that is less than 0.1×0.1 mm^2. Similarly, the number of defects also has a large difference, while the SD type has 9,505 samples, the WM type has only 36 samples. Besides, the contrast difference between the defects and background is low, such as SD and SR defect types. Moreover, MK and WM defect types are partially suppressed by the variant illumination of the smartphone's physical button surface. To overcome these challenges, we applied our HHoWST method to estimate the global illumination on the smartphone's physical button surface, thereby clarifying the error areas. Besides, we also used ASVDW to enhance the smartphone's physical button images in the DCL-CBI dataset. The result images by using HHoWST and ASVDW methods are shown in Fig. 14. As seen in Fig. 14(b), the compensated image of the two original images in Fig. 14(a) are enhanced after applying

ASVDW. However, since the ASVDW method is a global enhancement method, both background and defect regions are enhanced. On the other hand, the global illumination in the HHoWST images has been removed and the defect regions on the smartphone's physical button image have been retained and highlighted, as shown in Fig. 14(c).

The SSD is a compact CNN network and more suitable for detecting large objects, while it is not appropriate for detecting small objects. In this study, we used the SSD model to demonstrate the effectiveness of the proposed method in improving the detection rate for the small defect types, such as SD or PM defect types. The input image of the SSD network used in the experiment was resized to a size of 300×300 pixels and the learning rate was set to 0.001 during the training process. The backbone network used in this network is VGG16 [45], which was pre-trained by the ImageNet dataset [46] to reduce the convergence time of the training process. Compare to the previous YOLO versions, YOLOv5 is extremely faster and lightweight, while it can achieve the same accuracy as other benchmarks. The input image of the YOLOv5 model was set to a size of $1,280 \times 1.280$ pixels during the experiments. In contrast to the SSD model, Faster R-CNN uses the RPN network to predict the region containing the objects. Therefore, although this network costs much more time than SSD and YOLOv5 due to its larger architecture, it performs better in detecting small objects. In this study, we alternately used VGG16, ResNet101 [47], and MobileNet [48] as the backbone network of the Faster R-CNN model. While VGG16 and ResNet101 backbones perform better than the MobileNet backbone in extracting the feature of the image, it costs more time than using the MobileNet backbone, which is a small network and widely applied in real-time applications. Similar to the SSD network, the backbone networks used in Faster R-CNN were pre-trained on the ImageNet dataset to reduce the training time. The learning rate of Faster R-CNN was fixed at 0.001 during the training process. The CNN networks were trained by 100,000 iterations and monitored by checkpoints. Accordingly, the trained weights having the lower validation loss have been saved, the best weights were used for evaluating the test dataset. For training the DCL-CBI dataset with these deep learning models, we used 75% images for training and 25% remaining images for validation.

In this study, the mean average precision (mAP) is adopted as the primary metric to evaluate the effectiveness of recognition rate and determined as follows:

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N}, \qquad (33)$$

where N is the number of defect classes, while AP value is the average precision of the class i and calculated by the precision and recall rate as

$$AP = \frac{Precision + Recall}{2}, \qquad (34)$$

where *Precision* and *Recall* rates are determined as follows:

$$Precision = \frac{TP}{TP + FP}, \qquad (35)$$

$$Recall = \frac{TP}{TP + FN}, \qquad (36)$$

where TP, FP, TN, and FN denote the true positive, false positive, true negative, and false-negative rates, respectively.

The detection rates of the original images, ASVDW images and the HHoWST images by using different deep learning models with different backbones are illustrated in Table 2. Firstly, the effectiveness of the proposed method is evaluated by comparing the mAP values of the networks on the original images and HHoWST images. As seen, the mAP values of the network's training and evaluation on the HHoWST dataset are higher than the mAP values of the network's training and evaluation on the ASVDW images and original dataset. SSD is a compact network and it is not suitable for recognizing small objects. Therefore, SSD's recognition performance on the original images is only 85.9% for mAP. By using the SSD model, the mAP is significantly improved by 4.7% after applying our proposed method, while the improvement by using ASVDW is 0.8%. Similarly, by using the HHoWST method to highlight defect regions, the mAP of the YOLOv5 is increased by 1.3% from 83.7% to 85.0%. By using the ASVDW method to enhance images, the mAP of the YOLOv5 is improved 0.49% compared to the original images. The Faster R-CNN is a large network and it costs much more time during the training and evaluating process. However, the Faster R-CNN performs better than SSD and YOLOv5 networks in detecting the button defect. In particular, the mAP values of the Faster R-CNN model using VGG16 and ResNet101 as the backbone networks achieve 90.47% and 95.16%, respectively, for the original dataset. After applying ASVDW, the mAP of this model is increased by 0.65% and 0.01%, respectively. However, when applying Faster R-CNN model on HHoWST images, the mAP values are improved by 2.04% and 0.17%, for VGG16 and ResNet101 backbones, respectively. Although MobileNet is a small network, the mAP is also improved by 0.35% from 81.26 % to 81.61% on the HHoWST images compared to the original images, while the improvement is only increased by 0.13% when applying ASVDW to enhance the images.

One of the most challenges in defect detection is that the size of the defect is tiny. In the DCL-CBI dataset, most SD and PM defects are tiny. However, the number of SD and PM defects is large when compared to other defects. The average precision rates of this defect type in all experiments are lower than the average precision rates of other defects, lead to reducing the mean average precision rates of the models. The major cause of the lower average precision rates is that the performance of the deep learning models suffers the low recognition rate from the small size objects, especially the SSD model. By highlighting the defect region by using our proposed method, the average precision rates

TABLE 2. mAP and recognition rate (%) of the CNN networks evaluating on the original and preprocessed datasets.

NETWORK ARCHITECTURE	mAP	WM	CS	SR	PM	SD	MK
SSD300	85.90	100	99.80	74.90	72.70	68.30	99.60
SSD300 + ASVDW	86.70	100	100	77.51	75.40	70.10	97.20
SSD300 + HHoWST	90.60	100	100	82.30	89.50	80.90	90.80
YOLOv5	83.70	97.10	93.40	56.90	83.60	72.10	99.10
YOLOv5 + ASVDW	84.19	97.32	94.15	56.23	83.95	74.20	99.30
YOLOv5 + HHoWST	85.00	97.80	96.30	56.50	84.30	78.50	99.50
Faster R-CNN (VGG16)	90.47	100	98.69	80.25	86.78	77.11	100
Faster R-CNN (VGG16) + ASVDW	91.12	100	98.77	83.30	87.05	77.60	100
Faster R-CNN (VGG16) + HHoWST	92.51	100	98.87	88.59	89.13	78.47	100
Faster R-CNN (ResNet101)	95.16	100	99.89	90.10	90.85	90.09	100
Faster R-CNN (ResNet101) + ASVDW	95.17	100	99.70	90.21	90.87	90.22	100
Faster R-CNN (ResNet101) + HHoWST	95.33	100	99.78	90.87	90.88	90.43	100
Faster R-CNN (MobileNet)	81.26	95.80	98.14	76.33	64.69	61.66	90.91
Faster R-CNN (MobileNet) + ASVDW	81.39	96.74	98.30	75.20	65.36	61.98	90.78
Faster R-CNN (MobileNet) + HHoWST	81.61	99.30	98.61	70.64	67.06	63.50	90.57

of tiny defects, such as SD and PM defects, are improved significantly. In particular, the average precision rates of SD and PM by using the SSD model on HHoWST images are improved by 12.6% and 16.8% comparing to the original images, respectively. Similarly, the average precision rates in the YOLOv5 model for SD and PM defect types of the HHoWST images are also increased by 6.4% and 0.7%. The performance of the Faster R-CNN in detecting the tiny defects is better than the SSD and YOLOv5 models with VGG16 and ResNet101 as the backbone networks. However, the average precision rates of SD and PM defects by these models are still significantly lower than other defects. By using the Faster R-CNN with VGG16 backbone for the HHoWST images, the average precision rates of SD and PM defect types are improved by 1.36% and 2.35%, in comparison with the original images, respectively. Similarly, the average precision rates are increased by 0.34% and 0.03% for the Faster R-CNN using ResNet101 as the backbone network with our proposed method. The performance of the Faster R-CNN network using MobileNet as the backbone network in the detection of tiny defects is lower than other networks. The average precision rates of SD and PM defects are low and influence the mean average precision of the network. After using our proposed method to highlight the defect region, the average precision rates of SD and PM defects are improved by 1.84% and 2.37%, respectively. Therefore, after using the proposed method, the button defect is highlighted, which increases the overall recognition rate of the AOI system significantly. Especially, the tiny defect remains a challenge for any AOI system due to its less information about the size, shape, or color than other defects. The recognition rate of this defect

type can be improved after using our proposed method to enhance the feature.

To demonstrate more effectiveness of the proposed method in improving the recognition rate of an AOI system for tiny defects intuitively, the recognition result of the original images and the HHoWST images by the SSD, Faster R-CNN, and YOLOv5 model are further compared. Fig. 15, Fig. 16, and Fig. 17 present the smartphone's physical button defects detection results by using the SSD model with VGG16 backbone, Faster R-CNN model with VGG16 backbone, and YOLOv5 model, respectively. The detection results on the original images are shown in Fig. 15(b), Fig. 16(b), and Fig. 17(b), respectively. As seen, all three deep learning models cannot detect some tiny defects on the original images, which are indicated by the red arrows in Fig. 15(b), Fig. 16(b), and Fig. 17(b), respectively. However, all three deep learning models can detect effectively these tiny defects on the HHoWST images, as illustrated in Fig. 15(c), Fig. 16(c), and Fig. 17(c), respectively. Therefore, after preprocessing the smartphone's physical button images by the HHoWST method, deep learning models have good performance in detecting tiny defects, such as SD and PM defect types.

Among CNN networks used to evaluate the performance of the proposed method, the SSD network has the largest improvement in recognition rate than others when using our proposed method. Therefore, the dependence of the average precision on the intersection over the union (IoU) value of the button defect recognized by the SSD network is further discussed. The IoU is defined as the ratio of the overlap area over the union area between the predicted bounding box and
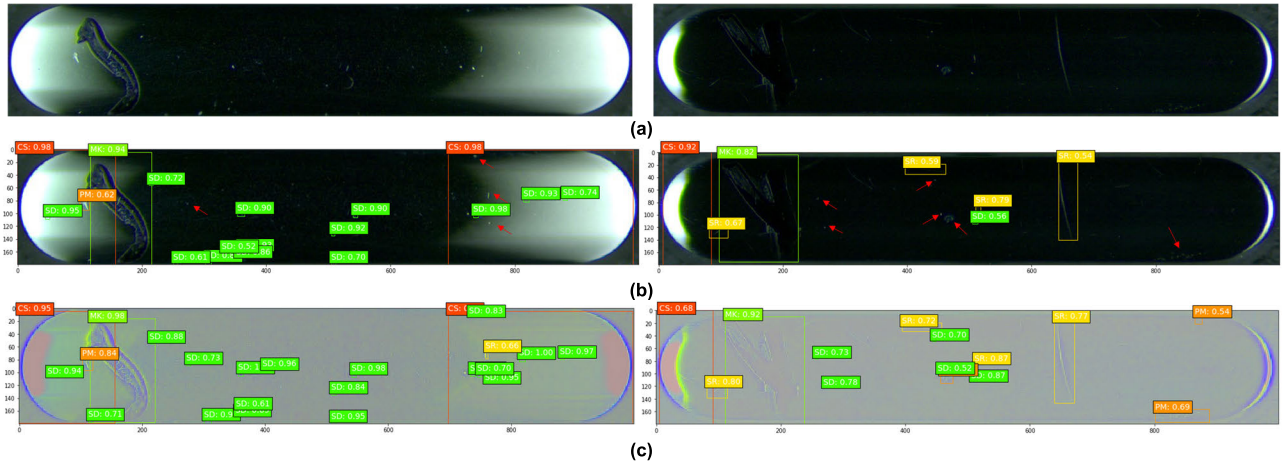
FIGURE 15. The defect detection results on the DCL-CBI dataset by the SSD model with VGG16 as the backbone network. (a) Two original images from the DCL-CBI dataset. (b) The detection results on the original images; (c) The detection results on the HHoWST images.
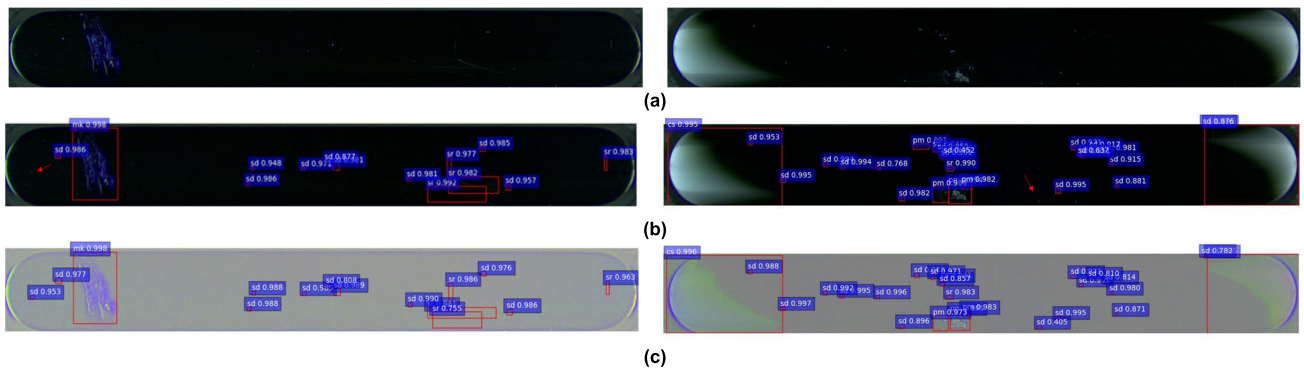
FIGURE 16. The defect detection results on the DCL-CBI dataset by the Faster R-CNN model with VGG16 as the backbone network. (a) Two original images from the DCL-CBI dataset. (b) The detection results on the original image; (c) The detection results on the HHoWST images.
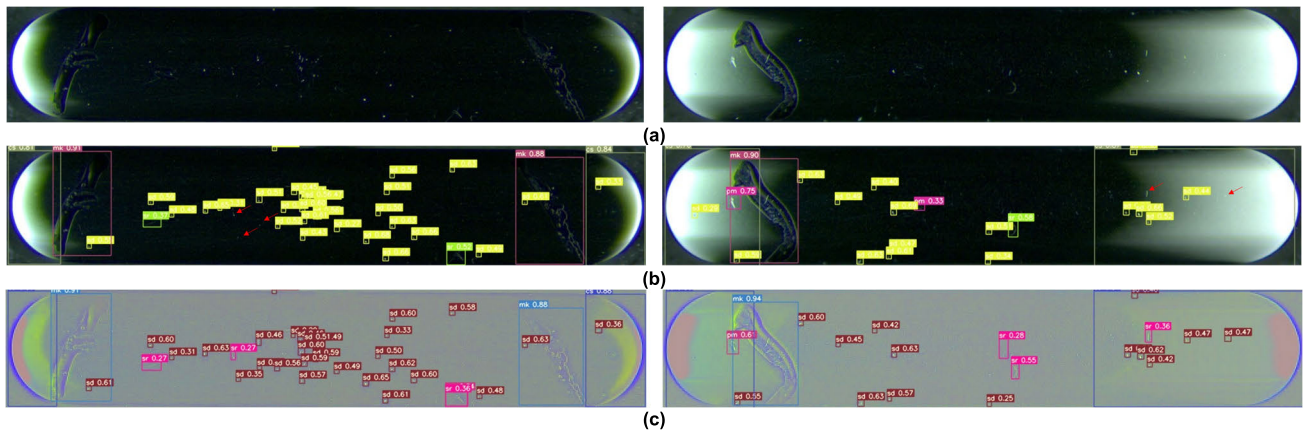
FIGURE 17. The defect detection results on the DCL-CBI dataset by the YOLOv5 model. (a) Two original images from the DCL-CBI dataset. (b) The detection results on the original images; (c) The detection results on the HHoWST images.

the ground-truth bounding box as follow

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}. \tag{37}$$

Therefore, the change of the IoU value is the main cause of the change in the recognition rates. Especially, the recognition rates of tiny defects are very sensitive to the change of the IoU value. In addition, using the SSD

network, which suffers many challenges from tiny defects, may increase the sensitivity of the recognition rate to the IoU value. Fig. 18 shows the changes in the average precision value evaluated on the original images (the solid line) and HHoWST images (the dotted line) upon the change of IoU value. Accordingly, the values of the average precision in the defect detection rate on HHoWST images are always larger than the average precision values evaluated on the

FIGURE 18. The dependence of the average precision on the IOU value of the button defects recognized by the SSD network.

original images. Especially, the difference between the average precision values of SD and PM defects evaluated on the original images and the HHoWST images is relatively large in the range of IoU of 0.26 - 0.62 for SD defect, and 0.35 - 0.7 for PM defect. Moreover, the average precision of WM defect of the original images decreases when the IoU value reaches 0.8, while HHoWST images still maintain the average precision of detect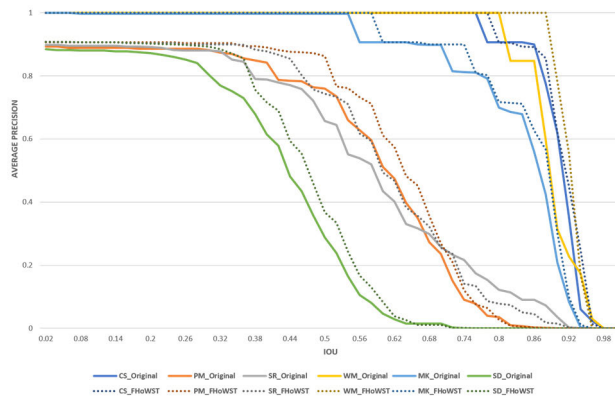ion rate at 100% until the IoU value reaches 0.88. Therefore, the increment of the IoU value reduces the average precision of the detection rate of the button defects, especially the tiny defects. Highlighting the defect regions by our proposed method can significantly reduce the effect of the IoU value on the recognition rate of deep learning models. In addition, compared to the small size defects, the defects with the larger size, such as WM or MK defects, are less impacted on the change of the IoU value when IoU is less than 0.8. However, once the IoU value reaches a larger value, the recognition rate of the button defect is decreased rapidly. The recognition rate of these defect types could be maintained with high values when the IoU value increases by using our proposed method to highlight defect regions.

V. CONCLUSION

This paper proposed a novel framework based on machine vision named HHoWST for real-time smartphone's physical buttons quality inspection. First, a modern image acquisition system is designed to obtain a high-quality smartphone's physical button image dataset with a total of 500 images, which contain 13,472 defect samples of six defect types. Next, the HHoWST method is proposed to estimate the global illumination and highlight the defective regions of the smartphone's physical button image. To do that, the wavelet subbands of three color channels of a smartphone's physical button image are used to constructs a third-order tensor, named WST, that contains both spatial information, correlations between color channels, and wavelet subbands of the color image. Through using higher-order singular value decomposition, we estimated the components containing the global illumination automatically, such as the coefficients of

the first frontal slice matrix of the core tensor and the 1st, 5th, 9th rows of inverse factor $U^{(3)}$. For the defect inspection task, we used three state-of-art deep learning models named SSD, Faster R-CNN, YOLOv5 to detect and classify the defects on the HHoWST images. Experimental results show that our proposed method significantly improves the defect detection efficiency of deep learning models, especially performance in detecting the tiny defect types when compared with the performance of defect detection on the original and ASVDW images.

REFERENCES

[1] R. T. Chin and C. A. Harlow, "Automated visual inspection: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, no. 6, pp. 557–573, Nov. 1982.

[2] A. A. R. M. A. Ebayyeh and A. Mousavi, "A review and analysis of automatic optical inspection and quality monitoring methods in electronics industry," *IEEE Access*, vol. 8, pp. 183192–183271, 2020.

[3] Z. Lu, Q. He, X. Xiang, and H. Liu, "Defect detection of PCB based on Bayes feature fusion," *J. Eng.*, vol. 2018, no. 16, pp. 1741–1745, Nov. 2018.

[4] F. Wu, X. Zhang, Y. Kuan, and Z. He, "An AOI algorithm for PCB based on feature extraction," in *Proc. 7th World Congr. Intell. Control Automat.*, 2008, pp. 240–247.

[5] B. Hu and J. Wang, "Detection of PCB surface defects with improved faster-RCNN and feature pyramid network," *IEEE Access*, vol. 8, pp. 108335–108345, 2020.

[6] W.-C. Wang, S.-L. Chen, L.-B. Chen, and W.-J. Chang, "A machine vision based automatic optical inspection system for measuring drilling quality of printed circuit boards," *IEEE Access*, vol. 5, pp. 10817–10833, 2017.

[7] V. V. Karlekar, M. S. Biradar, and K. B. Bhangale, "Fabric defect detection using wavelet filter," in *Proc. Int. Conf. Comput. Commun. Control Automat.*, Feb. 2015, pp. 712–715.

[8] Y. Li, W. Zhao, and J. Pan, "Deformable patterned fabric defect detection with Fisher criterion-based deep learning," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 1256–1264, Apr. 2017.

[9] C. Li, G. Gao, Z. Liu, M. Yu, and D. Huang, "Fabric defect detection based on biological vision modeling," *IEEE Access*, vol. 6, pp. 27659–27670, 2018.

[10] J. Liu, C. Wang, H. Su, B. Du, and D. Tao, "Multistage GAN for fabric defect detection," *IEEE Trans. Image Process.*, vol. 29, pp. 3388–3400, Dec. 2020.

[11] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1493–1504, Apr. 2020.

[12] S. Ghorai, A. Mukherjee, M. Gangadaran, and P. K. Dutta, "Automatic defect detection on hot-rolled flat steel products," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 3, pp. 612–621, Mar. 2013.

[13] K. Liu, N. Luo, A. Li, Y. Tian, H. Sajid, and H. Chen, "A new self-reference image decomposition algorithm for strip steel surface defect detection," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4732–4741, Jul. 2020.

[14] X. Cheng and J. Yu, "RetinaNet with difference channel attention and adaptively spatial feature fusion for steel surface defect detection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, Nov. 2021.

[15] X. Zhou, Y. Wang, C. Xiao, Q. Zhu, X. Lu, H. Zhang, J. Ge, and H. Zhao, "Automated visual inspection of glass bottle bottom with saliency detection and template matching," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 11, pp. 4253–4267, Nov. 2019.

[16] X. Zhou, Y. Wang, Q. Zhu, J. Mao, C. Xiao, X. Lu, and H. Zhang, "A surface defect detection framework for glass bottle bottom using visual attention model and wavelet transform," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2189–2201, Apr. 2020.

[17] X. Bi, C. Zhuang, and H. Ding, "A new Mura defect inspection way for TFT-LCD using level set method," *IEEE Signal Process. Lett.*, vol. 16, no. 4, pp. 311–314, Apr. 2009.

[18] S.-K. S. Fan and Y.-C. Chuang, "Automatic detection of Mura defect in TFT-LCD based on regression diagnostics," *Pattern Recognit. Lett.*, vol. 31, no. 15, pp. 2397–2404, 2010.

[19] A. Wu, J. Zhu, Z. Tao, and C. Mao, "Automatic inspection and classification for thin-film transistor liquid crystal display surface defects based on particle swarm optimization and one-class support vector machine," *Adv. Mech. Eng.*, vol. 8, no. 11, pp. 1–11, 2016.

[20] C.-J. Huang, C.-F. Wu, and C.-C. Wang, "Image processing techniques for wafer defect cluster identification," *IEEE Design Test Comput.*, vol. 19, no. 2, pp. 44–48, Apr. 2002.

[21] Y. Park and I. S. Kweon, "Ambiguous surface defect image classification of AMOLED displays in smartphones," *IEEE Trans. Ind. Informat.*, vol. 12, no. 2, pp. 597–607, Apr. 2016.

[22] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Trans. Comput.*, vol. C-23, no. 1, pp. 90–93, Jan. 1974.

[23] R. C. Gonzalez and R. E. Wood, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2008.

[24] S. Mallat, *A Wavelet Tour of Signal Processing. The Sparse Way*, 3rd ed. Amsterdam, The Netherlands: Elsevier, 2009.

[25] N. S. S. Mar, P. K. D. V. Yarlagadda, and C. Fookes, "Design and development of automatic visual inspection system for PCB manufacturing," *Robot. Comput.-Integr. Manuf.*, vol. 27, no. 5, pp. 949–962, Oct. 2011.

[26] H.-D. Lin and D.-C. Ho, "Detection of pinhole defects on chips and wafers using DCT enhancement in computer vision systems," *Int. J. Adv. Manuf. Technol.*, vol. 34, nos. 5–6, pp. 567–583, Aug. 2007.

[27] Y.-S. Chiu and H.-D. Lin, "An innovative blemish detection system for curved LED lenses," *Expert Syst. Appl.*, vol. 40, no. 2, pp. 471–479, Feb. 2013.

[28] D.-M. Tsai and C.-Y. Hung, "Automatic defect inspection of patterned thin film transistor-liquid crystal display (TFT-LCD) panels using one-dimensional Fourier reconstruction and wavelet decomposition," *Int. J. Prod. Res.*, vol. 43, no. 21, pp. 4589–4607, Nov. 2005.

[29] H.-D. Lin, "Automated defect inspection of light-emitting diode chips using neural network and statistical approaches," *Expert Syst. Appl.*, vol. 36, no. 1, pp. 219–226, Jan. 2009.

[30] A. S. Lewis and G. Knowles, "Image compression using the 2-D wavelet transform," *IEEE Trans. Image Process.*, vol. 1, no. 2, pp. 244–250, Apr. 1992.

[31] A. Averbuch, D. Lazar, and M. Israeli, "Image compression using wavelet transform and multiresolution decomposition," *IEEE Trans. Image Process.*, vol. 5, no. 1, pp. 4–15, Jan. 1996.

[32] J. W. Wang, N. T. Le, J. S. Lee, and C. C. Wang, "Illumination compensation for face recognition using adaptive singular value decomposition in the wavelet domain," *Inf. Sci.*, vol. 435, pp. 69–93, Apr. 2018.

[33] X. Wu, K. Xu, and J. Xu, "Application of undecimated wavelet transform to surface defect detection of hot rolled steel plates," in *Proc. Congr. Image Signal Process.*, vol. 4, May 2008, pp. 528–532.

[34] A. Laine and J. Fan, "Texture classification by wavelet packet signatures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 11, pp. 1186–1191, Nov. 1993.

[35] C. K. Chui, *Wavelet Analysis and its Applications*. San Diego, CA, USA: Elsevier, 1998.

[36] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics. 1992.

[37] W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus* (Series in Computational Mathematics). Leipzig, Germany: Springer, 2012.

[38] T. G. Kolda and B. W. Bader, "Tensor decomposition and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Sep. 2009.

[39] T. G. Kolda, "Multilinear operators for higher-order decompositions," Sandia Nat. Lab., Albuquerque, NM, USA, Tech. Rep. SAND2006-2081, 2006.

[40] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multi-modal factor analysis," *UCLA Work. Papers Phonet.*, vol. 16, pp. 1–84, Dec. 1970.

[41] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.

[42] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Cham, Switzerland: Springer, 2016, pp. 21–37.

[43] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[44] G. Jocher. *YOLOv5 in Pytorch*. [Online]. Available: https://github.com/ultralytics/yolov5

[45] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–14.

[46] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 248–255.

[48] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, Apr. 2017, pp. 1–9.

DUONG BINH GIAP received the B.S. degree in electronics and communications engineering from the Posts and Telecommunications Institute of Technology, Hanoi, Vietnam, in 2016, and the M.S. degree in electronics engineering from Le Quy Don Technical University, Hanoi, in 2018. He is currently pursuing the Ph.D. degree with the College of Electrical Engineering and Computer Science, National Kaohsiung University of Science and Technology, Taiwan. His research interests include digital image processing, pattern recognition, and face recognition.

TUYEN NGOC LE received the B.S. degree in mathematics and information from Hanoi National University of Education, Hanoi, Vietnam, in 1996, the M.S. degree in computer science from Le Quy Don Technical University, Hanoi, in 2006, and the Ph.D. degree in electronic engineering from the National Kaohsiung University of Science and Technology, Taiwan, in 2015. He is currently an Assistant Research Fellow with the Institute of Photonics Engineering, National Kaohsiung University of Science and Technology. His research interests include digital image processing, face recognition, fingerprint classification, and object detection.

JING-WEIN WANG (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from the National Taiwan University of Science and Technology, in 1986 and 1988, respectively, and the Ph.D. degree in electrical engineering from the National Cheng Kung University, Taiwan, in 1998. From 1992 to 2000, he was a Chief Project Leader with Equipment Design Center, Philips, Taiwan. He joined the Faculty of the National Kaohsiung University of Science and Technology, in 2000. From August 2017 to July 2019, he worked as the Dean of the College of Electrical Engineering and Information Science. He is currently a Distinguished Professor with the Institute of Photonics Engineering. His current research interests include combinatorial optimization, pattern recognition, and wavelets and their applications.

CHIA-NAN WANG received the Ph.D. degree from the Department of Industrial Engineering and Management, National Chiao Tung University, Taiwan, in 2004. He is currently a Professor with the Department of Industrial Engineering and Management, National Kaohsiung University of Science and Technology, Taiwan. He has published 80 journal articles and 40 patents. He achieves around 25 projects of research and academy and has some successful cases of technology transfer. His research interests include manufacturing automation and industrial applications. He also achieves awards from many international invention competitions (two Gold Medal Awards, five Silver Medal Awards, and several Bronze Medal Awards). In addition, he has plenty of practical experience in many industries and provides consultant services for several stock companies in Taiwan.

• • •