Optical Measurement of Highly Reflective Surfaces from a Single Exposure

Xingjian Liu, Wenyuan Chen, Harikrishnan Madhusudanan, Ji Ge, Changhai Ru, and Yu Sun

Abstract—Three-dimensional structured light (SL) measurement of highly reflective surface is a challenge faced in industrial metrology. The high dynamic range (HDR) technique provides a solution by fusing images under multiple exposures; however, the process is highly time-consuming. This paper reports a new SL-based method to measure parts with highly reflective surfaces from only a single exposure. A new quantitative metric is defined to optimally select camera exposure for capturing input single-exposure images. Different from existing image gradient or entropy-based metrics, the new metric incorporates both intensity modulation and overexposure. A skip pyramid context aggregation network (SP-CAN) is proposed to enhance the single exposure-captured images. Compared with existing image enhancement methods, SP-CAN effectively preserves detailed encoded phase information near edges and corners during enhancement. Experiments with various industrial parts demonstrated that the average time cost of the proposed method was 0.6 second, which was only one tenth of the HDR method (ten exposures), and the two methods achieved similar coverage rates (97.6% vs. 98.0%) and measurement accuracy (0.040 mm vs. 0.038 mm).

Index Terms—Industrial metrology, structured light (SL), reflective surface measurement, 3D point cloud, image enhancement, exposure selection.

I. INTRODUCTION

OPTICALLY measuring three-dimensional (3D) surface geometry has gained wide applications in industrial inspection [1], visual servoing [2], and vision-guided automation [3]. Among 3D optical measurement methods, structured light (SL)-based approach is popular due to its high-accuracy, non-contact, and full-field characteristics [4]. The SL-based

X. Liu, H. Madhusudanan, and J. Ge are with the Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON M5S 3G8, Canada, and also with the Robotics Institute, University of Toronto, Toronto, ON M5S 3G8, Canada. (e-mail: xj.liu@utoronto.ca; harikrishnan.madhusudanan@mail.utoronto.ca; geji1981@gmail.com).

W. Chen is with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G8, Canada, and also with the Robotics Institute, University of Toronto, Toronto, ON M5S 3G8, Canada. (e-mail: chenwy.chen@mail.utoronto.ca).

C. Ru is with the Research Center of Robotics and Micro System, Soochow University, Suzhou 215021, China, and also with the Jiangsu Provincial Key Laboratory of Advanced Robotics, Soochow University, Suzhou 215021, China. (e-mail: rzh@suda.edu.cn).

Y. Sun is with the Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON M5S 3G8, Canada, and also with the Robotics Institute, the Department of Electrical and Computer Engineering, and the Department of Computer Science, University of Toronto, Toronto, ON M5S 3G8, Canada. (e-mail: sun@mie.utoronto.ca).
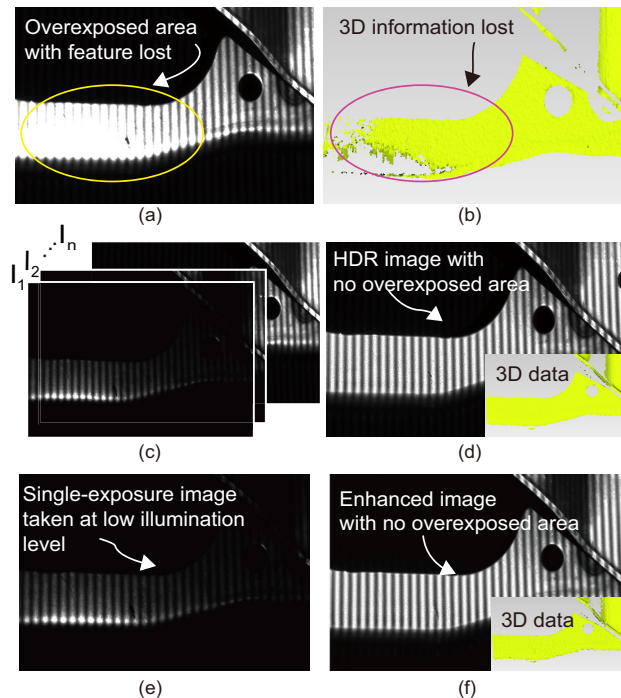

Fig. 1. (a) An automotive part with a large range of reflectivity variations taken under single exposure. (b) 3D reconstruction result from (a) by the traditional SL method with features lost in overexposed areas. (c) Traditional HDR method fusing multiple-exposure images. (d) Fused image of (c) and corresponding 3D reconstruction result. (e) Single-exposure image taken at a low illumination level. (f) Enhanced image of (e) and corresponding 3D reconstruction result.

method involves projecting a series of coded patterns to the target part. The surface of part is uniquely encoded by the projected patterns. Decoding the information from the captured images results in a 3D point cloud/mesh through triangulation [5], [6].

One challenge of the SL-based method is the measurement of parts that have highly reflective surfaces (e.g. metallic and electroplated surface) [7]. Fig. 1(a) shows a phase-coded image of an industrial part with a range of reflectivity variations. Fig. 1(b) shows the corresponding 3D reconstruction result by the traditional SL-based method, where only features under proper illumination can be recognized and well reconstructed while features in overexposed areas are lost. In practical applications, an effective way to deal with this problem is to spray white powder onto the target part to generate a diffuse surface [8]. However, powder spraying is time-consuming and changes surface properties of the target part, which is undesired in routine industrial applications. An alternative

solution is to place the light source or the camera at different viewpoints [9]. The limitation of this strategy is low measurement efficiency because the selection of viewpoints is empirical and computational cost of data fusion from different viewpoints is high [10].

For accurately measuring highly reflective surfaces using the SL-based method, the high dynamic range (HDR) technique was developed [11]–[13]. The HDR method involves measuring an part under multiple exposures. Features of different reflectance are captured under gradually increased exposures. Image fusion is then conducted to combine these features, as shown in Fig. 1(c)(d). Because the HDR method requires capturing and fusing images under multiple exposures (e.g., 5-20), the acquisition and fusion process is time consuming (3-10 seconds) [13].

This paper reports a new method for measuring highly reflective surfaces by using only a single exposure to realize rapid measurement while maintaining a high coverage rate and accuracy. In the proposed method, single-exposure images are taken at a relatively low illumination level to avoid information loss in overexposed areas, as shown in Fig. 1(e). Due to weak signals in single-exposure images taken under low illumination, the images are enhanced and used for subsequent 3D reconstruction (see Fig. 1(f)). Two challenges that must be tackled to realize the proposed single-exposure measurement include: (1) image enhancement must maintain the phase information in coded images during the enhancing process; and (2) the exposure time must be carefully selected to maximize the images intensity modulation while avoiding overexposure.

Traditional techniques for image enhancement, such as Wavelet Transform [14] and Retinex [15], enhance images for the purpose of visualization. They are based on the frequency domain or distribution of pixel intensities without consideration of encoded phase information. Therefore, the enhanced images fail to be decoded and reconstructed in the next step. Recently, convolutional neural network (CNN) has proven to be a powerful tool for solving image processing problems such as denoising [16], and restoration [17]. For such pixel-wise image enhancement task, state-of-the-art CNNs [18]–[20] use an atrous convolution [19], [20] to avoid the loss of high-frequency content caused by the traditional downsample-upsample structure [21]. However, existing atrous-convolution-based CNNs (e.g., context aggregation network (CAN) [19]) use a cascaded structure with only one large receptive field. Detailed features (e.g., deformed coding fringes in edge areas) covered by small receptive fields tend to be lost during image enhancement [22]. Although a pyramid structure (e.g., atrous spatial pyramid pooling (ASPP) [20]) has been developed to probe features from various size of receptive fields, it loses the advantage of detecting features with the large receptive field due to the practical limitation of network depth for avoiding overfitting [23].

For image enhancement, the quality of input images is of vital importance. In SL measurement, intensity modulation refers to the amplitude of coded patterns in captured images, and captured images with high intensity modulation are desired [5]. Increasing exposure time is a major physical

means to achieve increased intensity modulation [5], but long exposure time leads to irreversible information loss in overexposed areas. Therefore, an appropriate exposure time needs be selected in order to achieve a trade-off between preventing overexposure and maximizing intensity modulation of input images. Although such exposure time can be manually selected with trial and error based on visualization effect of images, visualization-based qualitative estimation is susceptible to subjective judgement [24]. Attempts have been made to select exposure time quantitatively for improving feature detection performance in visual perception tasks such as visual odometry and tracking [25]; however, the metrics in these methods are based on the maximization of image gradient [26] or image entropy [27] without considering two key aspects of SL images: intensity modulation and regional overexposure.

In this work, we designed a skip pyramid context aggregation network (SP-CAN) to enhance the single-exposure images while accurately maintaining encoded phase information. In the implementation of SP-CAN, features from intermediate layers covered by small receptive fields were integrated in a concatenation layer, which better preserves both small features (such as coded information near edges and corners) and large features (such as outlines of part) compared with the cascaded [19] or the pyramid structure [20]. Additionally, an optimal exposure selection technique was developed, in which a new metric was defined to incorporate both intensity modulation and overexposure. Experimental results with industrial parts having highly reflective surfaces demonstrated that our single-exposure technique achieved an average surface coverage rate 97.6% (vs. 57.3% by traditional SL [6]) and a low measurement error 0.040 mm (vs. 0.070 mm by traditional SL [6]) as benchmarked by a coordinate measuring machine (CMM). Compared with the HDR method [11], the proposed technique used only one tenth of its time while achieving a similar surface coverage rate and measurement accuracy.

II. SYSTEM OVERVIEW

A. System Setup

As shown in Fig. 2(a), the system consists of a blue LED projector (DLP LightCrafter 4500, Texas Instruments) with the resolution of 1140×912 pixels, and two Basler cameras (acA1440-73gm, resolution: 1456×1088). The focal length of both cameras is 12 mm. The working distance of the scanner is 700 mm, and the frame rate of the projector was set to 30 fps. In each measurement, twelve phase-coded images were projected onto a target part.

B. Operation Sequence

Fig. 2(b) illustrates the steps for highly reflective surface measurement with the proposed single-exposure method. For image acquisition, camera exposure is selected by an optimal exposure selection technique (Section III-B) to maximize intensity modulation while preventing overexposure. Since signals in acquired images are weak because of overexposure prevention, SP-CAN (Section III-A) is applied to amplify signals in the acquired images. After image enhancement, the encoded phase information within the images is decoded to
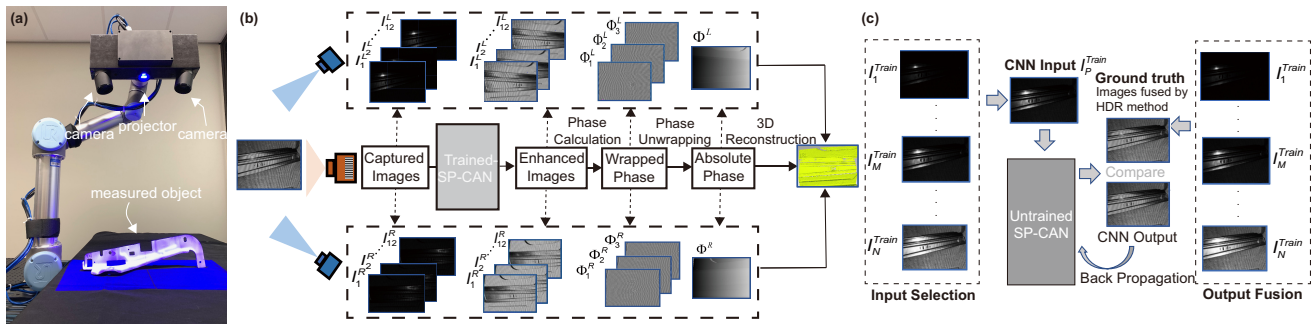
Fig. 2. (a) System setup: a custom-built structured-light scanner mounted on a robot arm. (b) Measurement process of the proposed method: image capturing, using trained SP-CAN to predict output images, using 4-step phase shift to calculate phase map, using multi-frequency heterodyne unwrapping to calculate the absolute phase map and final 3D reconstruction. (c) Training process of SP-CAN. The exposure time for capturing input images was chosen according to our proposed new metric, and output images were computed by the traditional HDR method.

recover the phase map. Finally, a point cloud is reconstructed based on triangulation.

C. Image Dataset and Training Process

For training and evaluating the performance of SP-CAN, a dataset was constructed in this work from ten different automotive parts (e.g., front end stamping parts, cross car beam bracket, fuel filler door, etc.). The Highly-Reflective-Dataset, available for download at https://github.com/chenwy960424/Highly-Reflective-Dataset, contains 900 raw single-exposure images, each with a corresponding HDR-computed reference image. The exposure time for capturing these raw single-exposure images was selected by the proposed optimal exposure selection technique. The slow but highly accurate HDR technique [11] was used to compute reference images. The starting and ending exposure used in the HDR method was set to be 30 ms and 300 ms with an interval of 30 ms.

From the dataset, 20% of the images were randomly selected as the test set, and the rest 80% were used as the training set. During the training process, the raw single-exposure images were used as input while the HDR-computed images were compared with output images of SP-CAN, as shown in Fig. 2(c). The weights of the network were initialized by the identity initialization technique [19]. Then the Adam optimizer together with backpropagation was used to update the SP-CAN weights. Meanwhile, an early stopping rule was imposed, i.e., the training procedure is stopped if training error does not decrease after ten epochs.

III. METHODOLOGY

To achieve highly reflective surfaces measurement using a single exposure, two key techniques were developed in this work. SP-CAN was designed to enhance input single-exposure images while retaining encoded phase information; and an optimal exposure selection technique was developed to select a proper camera exposure time for capturing input images.

A. Skip Pyramid Context Aggregation Network (SP-CAN)

The enhancement of input single-exposure images is in essence a pixel-wise recovery problem. Traditional techniques such as Wavelet Transform [14] and Retinex [15] are mostly for visualization enhancement. They do not maintain encoded phase information within captured images. Among recent advances in learning-based image processing techniques [19], convolutional neural network (CNN) solves the enhancement problem by building a statistical model. To maintain encoded phase information within adjacent pixels during image enhancement, the calculation of each pixels value should include its surrounding pixels. The area of surrounding pixels contributing to the calculation is referred to as the receptive field, and collecting information from receptive fields of various sizes is necessary for building the statistical model.

A popular approach to change the size of a receptive field in CNN-based methods is downsampling and upsampling of the input image [16], [17]. Due to the change of image resolution during the downsampling and upsampling process, such methods suffer from the loss of high-frequency contents [21]. Atrous convolution [19], [20] was thus proposed to effectively change the receptive field without sacrificing image resolution and high-frequency contents. The atrous convolution $*_r$ is generated by introducing $r-1$ zeros between two consecutive convolutional kernels in the traditional convolution, where r is termed atrous rate (see Fig. 3(b)). For an $m \times n$ atrous convolution kernel, the corresponding size of the receptive field is $(mr+r-1) \times (nr+r-1)$ which grows with the atrous rate r. Therefore, the atrous rate controls the size of the receptive field.

As shown in Fig. 3(a), existing atrous convolution-based CNNs such as CAN [19] stack atrous convolutional layers with increasing atrous rates in a cascaded fashion. Cascading these convolutional layers produces a large receptive field. For example, suppose there are L layers in CAN, and each layer has the atrous rate of $r_i, i = 1, 2, \ldots, L$. After cascading these L convolutional layers, the kernel size R_{sum} of the receptive field ($R_{sum} \times R_{sum}$) is

$$\begin{cases} R_{sum} = \sum_{i=1}^{L} R_i - (L-1) \\ R_i = (Kr_i + r_i - 1) \end{cases} \quad (1)$$

where R_i is the size of the receptive field of the ith layer, and K represents the convolution kernel's size (for the example
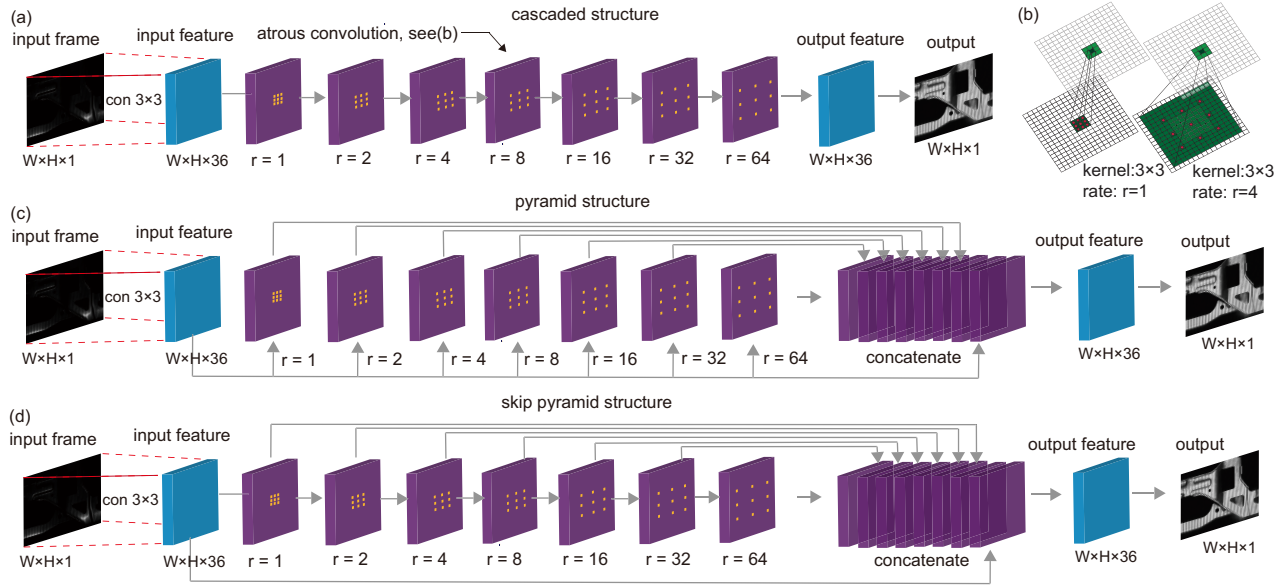
Fig. 3. (a) Structure of a traditional atrous convolution-based CNN, context aggregation network (CAN) [19]. CAN has a cascaded structure and after a series of atrous convolutions, in the output of the last layer, output features have a single large receptive field, and detailed features covered by small receptive fields are lost. (b) Atrous convolution kernel with atrous rate $r = 1$, receptive field $R = 3 \times 3$ (left). Atrous convolution kernel with atrous rate $r = 4$ (three zeros between two adjacent kernels), receptive field $R = 15 \times 15$ (right). (c) Atrous convolution network with atrous spatial pyramid pooling (ASPP) has a pyramid structure [20] for probing features with receptive fields of various sizes. However, it loses the advantage of detecting features with a large receptive field. (d) Our proposed skip pyramid context aggregation network (SP-CAN). To preserve both advantages of cascaded structure and pyramid structure, SP-CAN utilizes skip connections to extract the outputs of each layer and combines them in a concatenation layer to generate final prediction.

shown in Fig. 3(a), $K = 3$). Therefore, all extracted features in CAN share the same large receptive field R_{sum}. CAN was originally designed to detect large objects such as horses and people where a large receptive field is needed. However, in our work, detailed phase information near corners and edges covered by small receptive fields are also required. The single large receptive field in CAN neglects such information and causes recovery errors, as further shown in the Results section.

To overcome this limitation, an atrous convolutional network with a pyramid structure [20] was developed. In the pyramid structure, atrous convolution layers with increasing atrous rates are organized in a parallel fashion (see Fig. 3(c)). This design aimed at probing features with receptive fields of various sizes. However, due to the practical limitation of the network depth in the pyramid structure for avoiding overfitting, the pyramid structure loses the advantage of detecting features with a large receptive field, such as outline of an object. To preserve both advantages of the cascaded structure and the pyramid structure, we propose a skip pyramid context aggregation network (SP-CAN) by adding skip connections to integrate features covered by both small and large receptive fields. A skip connection directly feeds the output of one layer as the input for another non-adjacent layer by skipping all other layers in between [28]. Since the atrous rates of intermediate layers are different, the corresponding receptive fields of these layers have various sizes covering features from small to large. As shown in Fig. 3(d), the skip connections pass features extracted by intermediate layers to a concatenation layer, where features from receptive fields of different sizes are fused.

In SP-CAN, L atrous convolution layers with incremental

atrous rates are firstly cascaded, according to

$$
\begin{cases}
F_m^l = \Phi(\Psi(b_m^l + \sum_{n=1}^{N} F_n^{l-1} *_{r_l} K_{m,n}^l)) \\
\Psi(x) = \lambda x + \mu BN(x) \\
\Phi(x) = \max(0.2x, x)
\end{cases} \quad (2)
$$

where F_m^l represents the mth feature map of lth layer, F_n^{l-1} is the nth feature map of the previous layer, $K_{m,n}^l$ is a 3×3 convolution, and b_m^l is a bias term. Ψ is the adaptive normalization for preprocessing data, where BN is the batch normalization function, and λ and μ are coefficients that are automatically learned during training. $\Phi(x)$ is a pointwise nonlinear function. The operator $*_{r_l}$ denotes the atrous convolution with the rate of $r_l = 2^{l-1}$. The proposed SP-CAN utilizes skip connections $\mathbb{S}(F)$ to extract the outputs of each layer F_m^l and combines them in the concatenation layer, according to

$$
F_o = \Phi(\Psi(b_s^l + \sum_{l=1}^{L} \sum_{m=1}^{M} \mathbb{S}(F_m^l) * K_{m,n}^l)) \quad (3)
$$

where the output feature map F_o covers a large range size of receptive fields.

Finally, since our SP-CAN is designed for image enhancement, which is a regression problem rather than a classification problem, the widely used softmax layer [29] is replaced with the mean squared error (MSE) loss function to produce continuous-valued intensity for each pixel. MSE is computed by comparing a generated image and the reference image.

$$
MSE = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} \left\| INT_o(x,y) - INT_g(x,y) \right\|^2 \quad (4)
$$

where W, H are the image width and height, $INT_o(x, y)$ is the intensity for pixel (x,y) in the output image, and $INT_g(x, y)$ is the intensity in the reference image. The SP-CAN effectively enhances images while maintaining detailed encoded phase information near edges and corners.
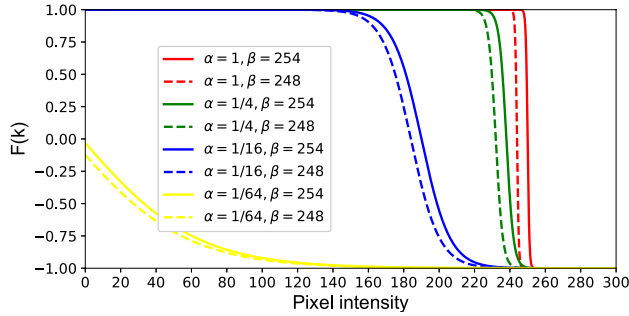
B. Optimal Exposure Selection

Fig. 4. Activation function that controls the weight of a pixel based on its intensity. The function or weight is influenced by α and β, where α determines sensitivity to overexposure and β determines threshold of overexposure.

In SL measurement, intensity modulation refers to the amplitude of coded patterns in the captured images, and images with high intensity modulation are desired [5]. Increasing exposure is a common approach to improve intensity modulation; however, overexposure can lead to irreversible information loss. For quantitative exposure selection, existing metrics are based on the maximization of gradient [26] or entropy [27] for the purpose of improving feature detection or visual effect in visual perception tasks such as visual odometry and tracking [25]. Since intensity modulation and overexposure are the two key aspects characterizing the quality of SL images and neither gradient-based methods [26] nor entropy-based methods [27] accounted for intensity modulation and overexposure, we propose a new metric which incorporates both intensity modulation and overexposure for the task of measuring highly reflective surfaces.

Based on the gradient and entropy-based metrics [26] [27], substituting gradient and entropy with intensity modulation gives

$$\mathbf{Q} = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} M(x, y) \qquad (5)$$

The intensity of each pixel $INT(x, y)$ in the acquired image is decomposed into two parts, including intensity modulation $M(x, y)$ and intensity of ambient light $A(x, y)$, i.e.,

$$\begin{cases} INT(x, y) = M(x, y) + A(x, y) \\ M(x, y) = fta(x, y)L_p \\ A(x, y) = ftb(x, y) \end{cases} \qquad (6)$$

where t is the camera exposure, f is the coefficient of camera sensitivity, a represents surface reflectivity, L_p is the intensity of projected light, and b is the intensity of ambient light. By capturing images, image I_1 taken without projected light, and

image I_2 taken with a projected uniform pattern, then $M(x, y) = I_2(x, y) - I_1(x, y)$.

Note that the metric in (5) neglects the information loss caused by overexposure. For example, suppose an area $O_2 \subseteq I_2$ is overexposed within image I_2 taken with a projected uniform pattern, and the corresponding area in I_1 is $O_1 \subseteq I_1$. Because of overexposure, the intensity of the pixels is saturated at 255 in O_2, which makes the modulation value in O_2 spuriously high, and coding information is irreversibly lost in O_2. Thus, in order to suppress overexposure, an activation function $\mathbf{F}(k), k \in [0, 255]$ is added to (5) to constitute the metric,

$$\mathbf{Q}_m = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} \mathbf{F}(I(x, y)) M(x, y) \qquad (7)$$

with

$$\mathbf{F}(k) = \frac{exp(-\alpha(k - \beta) - 4) - exp(\alpha(k - \beta) + 4)}{exp(\alpha(k - \beta) + 4) + exp(-\alpha(k - \beta) - 4)} \qquad (8)$$

The activation function $\mathbf{F}(k)$ is a variant of the classic tanh function [30]. It controls overexposure by assigning smaller weights to close-to-overexposed pixels. $\alpha \in (0, 1]$ determines the balance between intensity modulation and overexposure. The larger the α, the faster the weight decreases when the pixel is considered as overexposed, as shown in Fig. 4. Since overexposure control is critical for the task of highly reflective surface measurement, α is set to be 1. The value of $\beta \in (0, 255)$ determines the threshold of overexposure. Pixels with an intensity value in $(\beta, 255]$ are regarded as overexposed and suppressed in the metric (7). In this work, we set $\beta = 248$.

The proposed metric (7) is used to measure the quality of input images. To maximize the metric value, its partial

---

Algorithm 1 Optimal exposure selection.

Input: Pre-calculated system parameters: f, a, b, L_p; Initial query exposure: $t_{init} = 0.5 * max\_exposure$; Pre-defined parameters: $THR = 10^{-5}, \gamma = 5 \times 10^5, MAX\_ITER = 10$;
Output: Optimal exposure t_o;
1: Initialization: $t = t_{init}$;
2: **while** $i < MAX\_ITER$ **do**
3:      Under exposure t, capture image I_1 without projected light, capture image I_2 with projected light;
4:      Calculate intensity modulation $M(x, y)$ with I_1 and I_2;
5:      Calculate the evaluation metric $\mathbf{Q}_m \leftarrow$ (7);
6:      Calculate metric's gradient $\partial \mathbf{Q}_m \leftarrow$ (9) ;
7:      **if** $|\partial \mathbf{Q}_m| < THR$ **then**
8:          //the metric value is considered as the highest when the gradient reaches convergence
9:          **return** the output: $t_o = t$;
         break;
10:      **else**
11:          Update the exposure t using (11);
12:      **end if**
13: **end while**
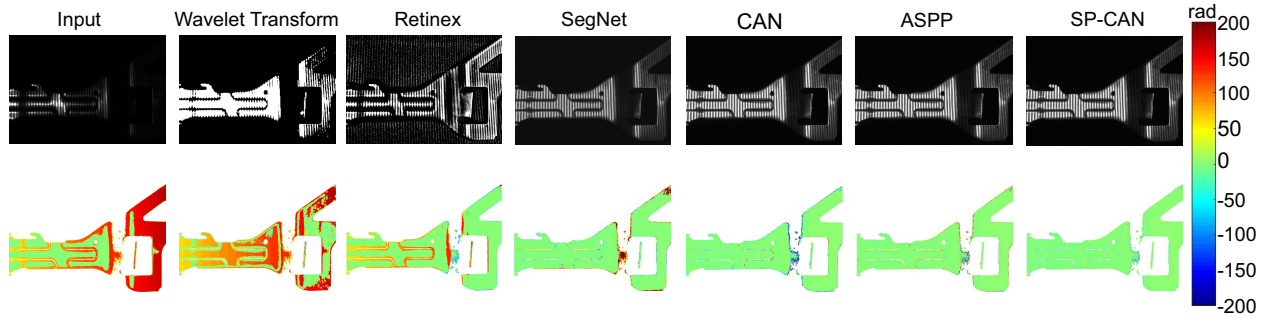14: **return** Optimal exposure result t_o;

Fig. 5. Comparison of different image enhancement methods: Wavelet Transform (WT) [14], Retinex [15], SegNet (downsample-upsample structure) [18], CAN (cascaded structure) [19], ASPP (pyramid structure) [20], and our proposed SP-CAN. For each method, shown is image after enhancement as well as the corresponding per-pixel phase error map, using ten-exposure HDR as benchmark.

derivative with respect to exposure time is

$$\frac{\partial \mathbf{Q}_m}{\partial t} = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} \left(\frac{\partial M(x,y)}{\partial t} \mathbf{F}(I(x,y)) + \frac{\partial \mathbf{F}(I(x,y))}{\partial t} M(x,y) \right) \tag{9}$$

with

$$\begin{cases} \dfrac{\partial \mathbf{F}(I(x,y))}{\partial t} = \dfrac{\partial \mathbf{F}(I(x,y))}{\partial I(x,y)} \dfrac{\partial I(x,y)}{\partial t} \\[2mm] \dfrac{\partial M(x,y)}{\partial t} = f\alpha(x,y)L_p \\[2mm] \dfrac{\partial I(x,y)}{\partial t} = f\alpha(x,y)L_p + fb(x,y) \end{cases} \tag{10}$$

The parameters f, a, b, L_p are fixed for a given SL measurement system [31]. Algorithm 1 describes the entire working flow. In gradient descent, the initial exposure is set to be the mid-value of maximum exposure. THR denotes the threshold value for the gradient of the proposed metric, $|\partial \mathbf{Q}_m|$ for determining the convergence of $\mathbf{Q}_m$. MAX_ITER denotes means the maximum number of iterations in searching for the extremum of $\mathbf{Q}_m$ to prevent the algorithm from being trapped in a dead loop. It avoids getting stuck in a dead loop. The threshold values for THR and MAX_ITER are tuned experimentally, using the phase means absolute error (PMAE) shown in (12) for quantitative evaluation. The coefficient of gradient descent γ is set to 5×10^5. The exposure time is adjusted from the latest exposure as

$$t = t + \gamma \frac{\partial \mathbf{Q}_m}{\partial t}. \tag{11}$$

IV. EXPERIMENTAL RESULTS

A. Performance of Different Image Enhancement Methods

To evaluate the performance of different image enhancement methods, phase means absolute error ($PMAE$) was used [32].

$$PMAE = \frac{1}{W \times H} \sum_{x=1}^{W} \sum_{y=1}^{H} |P_o(x,y) - P_b(x,y)| \tag{12}$$

where W, H are the width and height of the image, and $P_o(x,y)$ is the phase value for pixel (x,y) in the enhanced image, and $P_b(x,y)$ is the phase value for pixel (x,y) in benchmark phase map. The smaller the $PMAE$ value is, the higher the accuracy is for the calculated phase map.

TABLE I
PERFORMANCE OF IMAGE ENHANCEMENT BY DIFFERENT METHODS

	Input	WT	Retinex	SegNet	CAN	ASPP	SP-CAN
$PMAE$ (rad)	85.1	75.5	21.2	10.2	7.1	5.8	0.85

In the experiment, a series of phase-coded images were captured under a low exposure value (with no overexposed region) and used as input single-exposure images I_{input} for image enhancement. In the meanwhile, to obtain the benchmark phase map, phase-coded images were repeatedly captured with ten gradually increased exposures, and the benchmark P_b was obtained by the traditional HDR method [11]. The performance of our proposed SP-CAN was compared with that of traditional enhancement methods including Wavelet Transform [14] and Retinex [15], and CNN-based methods including SegNet (downsample-upsample structure) [18], CAN (cascaded structure) [19] and ASPP (pyramid structure) [20]. These methods were applied to enhance the same input single-exposure images I_{input}, and then the conventional four-step phase shifting algorithm [6] was used to calculate the phase maps P_o. Note that all the CNN-based methods were retrained by using the same dataset as SP-CAN. By comparing to the benchmark P_b, $PAME$ of each method was quantified. The enhanced images and corresponding phase error map are shown in Fig. 5, and their respective $PMAE$ values are summarized in Table I.

As shown in Fig. 5, although traditional methods (Wavelet Transform [14] and Retinex [15]) well enhanced the input image for visualization, the encoded phase information was lost. The corresponding $PMAE$ values were as high as 75.5 and 21.2 rads (see Table. I), which were only slightly better than the single-exposure method without enhancement (85.1 rads). This was due to the omission of encoded phase information during image enhancement in these traditional methods. Among all the CNN-based methods, the accuracy of SegNet was the poorest with a $PMAE$ value of 10.2 rads, due to the loss of high-frequency contents caused by downsampling and upsampling [21]. In comparison, astrous-convolution-based methods showed better performance (7.1 rads for CAN and 5.8 rads for ASPP). However, detailed information near edges and corners was lost in CAN while large features such as the object's outlines were lost in ASPP.

This was because all extracted features in CAN [19] share the same large receptive field caused by the cascaded structure while such features covered by the large receptive field was not covered in the pyramid structure. Different from these two networks, the skip connections and the concatenation layer in our proposed SP-CAN integrated features from small to large and effectively preserved both detailed and outlined encoded phase information, resulting in the lowest *PAME* value of 0.85 rad.

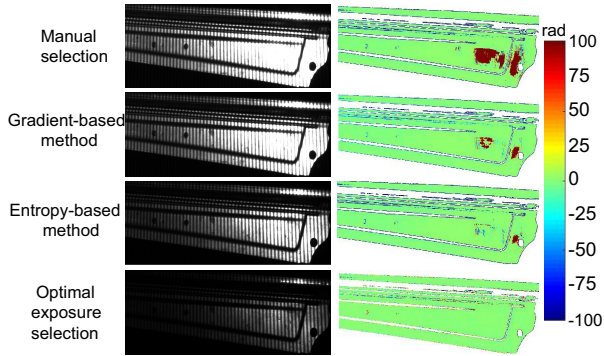### B. Performance of Optimal Exposure Selection

Fig. 6. Comparison of different exposure selection methods: manual selection, gradient-based method, entropy-based method, and our proposed method. For each method, shown is the image taken under the selected exposure as well as the corresponding per-pixel phase error map (all enhanced by SP-CAN).

The quality of input single-exposure images greatly affects the image enhancement result and subsequent 3D reconstruction. Since the goal for improving the quality of input images is to boost the performance of image enhancement, the phase means absolute error (*PMAE*) was again used to evaluate the result of exposure selection. The HDR method with ten exposures was also used for benchmark.

We experimentally compared our proposed method, gradient-based method [26], entropy-based method [27], and manual selection. A set of phase-coded images were first captured under the exposure selected by each method and then all enhanced by SP-CAN. After calculating the phase map for enhanced images, the *PMAE* value for each method was determined by comparing with the HDR-obtained benchmark data. The input images captured with the exposure time selected by each method as well as their corresponding phase error maps are shown in Fig. 6, and their selected exposures and *PMAE* values were summarized in Table II. Among these methods, the *PMAE* value of manual selection was the poorest (12.48 rads) because manual selection was based on the visualization effect. In comparison, other three methods are all based on quantitative selection and all produced better results with *PMAE* values below 12 rads. In particular, the proposed method outperformed the gradient-based [26] and entropy-based [27] methods (0.91 rad vs 11.5 rads; 0.91 rad vs 8.6 rads) because the proposed method maximizes intensity modulation and better preserves the encoded phase information for image enhancement. Additionally, the gradient-based and entropy-based methods permitted a higher level of overexposure for increasing gradient or entropy; therefore, they suffered from

irreversible information loss caused by overexposure (see Fig. 6). In comparison, the activation function in our proposed method effectively prevented overexposure.

TABLE II
COMPARISON OF EXPOSURE SELECTION METHODS

Symbol	Manual Selection	Gradient-based	Entropy-based	Proposed method
Exposure (ms)	250.0	228.4	121.7	43.2
PMAE (rad)	12.48	11.50	8.60	0.91

C. Evaluation of Point Cloud Coverage Rate

To evaluate the performance of measuring highly reflective surfaces, three different industrial parts with different reflectivity were measured (see Fig. 7, first three rows). The coverage rate Υ, used for quantitative evaluation, is defined as

$$\Upsilon = \frac{S_m}{S_b} \times 100\% \qquad (13)$$

where S_m is the surface area of the measured point cloud, and S_b represents the surface area calculated from benchmark data. In this experiment, the benchmark point cloud was obtained by using the HDR method. For each part, camera exposure was selected by the proposed optimal exposure selection technique, and then the input single-exposure images were enhanced by SP-CAN. The enhanced images were used to calculate the point cloud, and the coverage rate was quantified via comparing with benchmark data. The performance of the proposed method was compared with that of single-exposure without enhancement, HDR method with three exposures, and HDR method with ten exposures [11].

As shown in Fig. 7 and Fig. 8 (a), both qualitative and quantitative comparisons were made, which confirming the effectiveness of overexposure reduction by our proposed method. As shown in Fig. 7, the proposed method can better recover overexposed areas than the method of single-exposure without enhancement (average surface coverage rate: 97.6% vs. 57.3%). Although HDR with three exposures was capable of partially recovering the lost features compared to the single-exposure method, its average surface coverage rate was much lower than that measured by our proposed method (90.6% vs. 97.6%). HDR with ten exposures (98.0%) slightly outperformed our proposed method by 0.4% in average surface coverage rate; however, the time cost of HDR with ten exposures was more than ten times that of our proposed method (Fig. 8(b)). Thus, our proposed method (SP-CAN + optimal exposure selection) was capable of achieving similar surface coverage rates as ten-exposure HDR but with one tenth time cost.

To prove the generalization ability of our proposed SP-CAN, it was used to recover 3D information of a new industrial part (fuel filler door) that SP-CAN had not seen. The surface reflectivity of the fuel filler door was different from that of other parts in the original dataset. The camera exposure was selected by the proposed optimal exposure selection method, and the input images were enhanced by SP-CAN. The qualitative (Fig. 7 last row) and quantitative (Fig. 8(a)) results showed that our proposed SP-CAN + optimal exposure

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2020.2991458, IEEE Transactions on Industrial Informatics
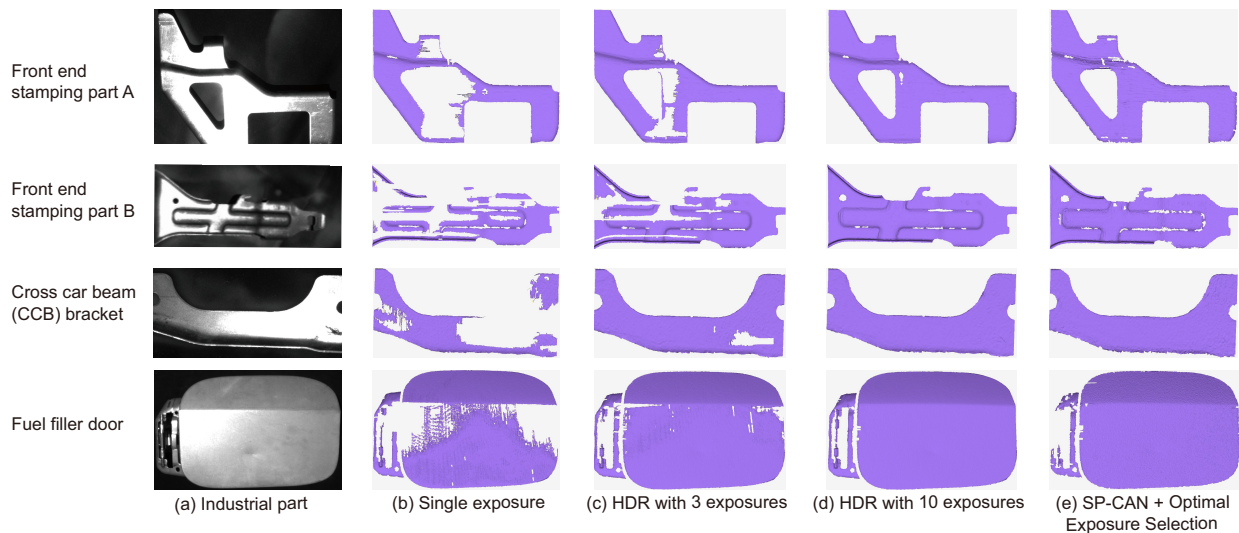
8

Fig. 7. 3D reconstruction of industrial parts. For each part, shown are original image and 3D reconstruction result measured by different methods: single-exposure (without enhancement), HDR with three exposures, HDR with ten exposures, and proposed SP-CAN + optimal exposure selection. The fourth row shows the results from fuel filler door that was never seen by SP-CAN to prove its generalization ability.

selection method achieved a similar (slightly lower) surface coverage for the newly added sample as the original training samples (95.2% vs. 97.6%).
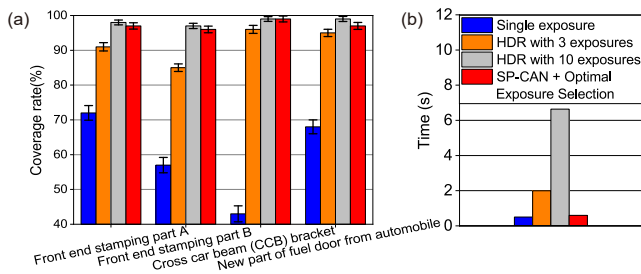

Fig. 8. Comparison of (a) surface coverage rate and (b) time cost of measuring different industrial parts by, the single-exposure method (without enhancement), HDR with three exposures, HDR with ten exposures, and our proposed SP-CAN + optimal exposure selection. The fourth column in (a) shows the results from a fuel filler door had not been seen by SP-CAN to prove its generalization ability.

D. Point Cloud Accuracy Evaluation

To evaluate the measurement accuracy of the proposed method, a standard ball-bar shown in Fig. 9(a) verified by a coordinate measuring machine (CMM) was used as the part to measure. It consists of two ceramic spheres with the diameter of 38.1043 mm and 38.1048 mm and center-to-center distance of 100.2072 mm. The absolute measurement error $\varepsilon = |L_m - L_b|$ was used as the evaluation metric, where L_m is the measured value and L_b is the CMM-measured data.

After the 3D point cloud of each sphere was acquired, a best-fit sphere was calculated with the least-squares method, as shown in Fig. 9(b). Then the distance between centers of the two spheres was measured. The measurement accuracy was evaluated by calculating the error ε for measured center-to-center distance. For repeatability, the measurement was conducted ten times from different positions and orientations.

The result of the proposed method was compared with that of single-exposure without enhancement and HDR with ten exposures. As summarized in Fig. 9(c), the performance of single exposure without enhancement was poor, with a mean absolute error $\varepsilon = 0.070mm$ for the measured center-to-center distance due to the loss of information in overexposed regions. In comparison, the proposed method and the ten-exposure HDR method both achieved much higher accuracy ($\varepsilon = 0.040mm$ and $\varepsilon = 0.038mm$).
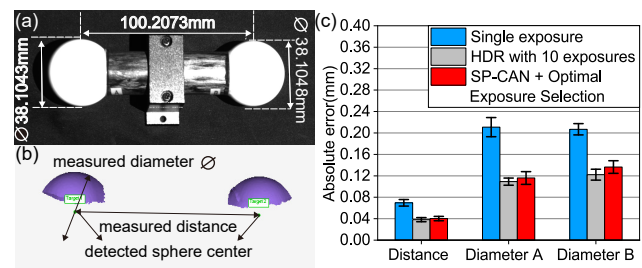

Fig. 9. (a) Standard ball-bar used for evaluating measurement accuracy. (b) Captured 3D point cloud. (c) Comparison of absolute error for single exposure without enhancement, HDR with ten exposures, and our proposed method in terms of the measured distance between two centers and the balls' diameters.

V. CONCLUSION

This paper presented a new technique for SL measurement of parts with highly reflective surfaces. Single-exposure images were effectively enhanced by our proposed SP-CAN method, and exposure time selection was quantitatively determined. By measuring various parts with highly reflective surfaces, the experimental results revealed that the proposed method achieved a surface coverage rate of 97.6% and an absolute measurement error of 0.040 mm. Compared with traditional ten-exposure HDR, our proposed method achieved similar accuracy and coverage rate but the time cost was only one tenth.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2020.2991458, IEEE Transactions on Industrial Informatics

9

REFERENCES

[1] S. von Enzberg and A. Al-Hamadi, "A multiresolution approach to model-based 3-d surface quality inspection," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 4, pp. 1498–1507, 2016.

[2] A. Anwar *et al.*, "Quality inspection of remote radio units using depth-free image-based visual servo with acceleration command," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 10, pp. 8214–8223, 2018.

[3] W. Lin *et al.*, "Recognition and pose estimation of auto parts for an autonomous spray painting robot," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1709–1719, 2018.

[4] Z. Xiong *et al.*, "Computational depth sensing: Toward high-performance commodity depth cameras," *IEEE Signal Processing Magazine*, vol. 34, no. 3, pp. 55–68, 2017.

[5] J. Geng, "Structured-light 3d surface imaging: a tutorial," *Advances in Optics and Photonics*, vol. 3, no. 2, pp. 128–160, 2011.

[6] Z. Li *et al.*, "Multiview phase shifting: a full-resolution and high-speed 3d measurement framework for arbitrary shape dynamic objects," *Optics letters*, vol. 38, no. 9, pp. 1389–1391, 2013.

[7] Z. Song *et al.*, "An accurate and robust strip-edge-based structured light means for shiny surface micromeasurement in 3-d," *IEEE Transactions on Industrial Electronics*, vol. 3, no. 60, pp. 1023–1032, 2013.

[8] D. Palousek *et al.*, "Effect of matte coating on 3d optical measurement accuracy," *Optical Materials*, vol. 40, pp. 1–9, 2015.

[9] C. Hong *et al.*, "Image-based three-dimensional human pose recovery by multiview locality-sensitive sparse retrieval," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 6, pp. 3742–3751, 2014.

[10] S. Wei *et al.*, "Colorful 3-d imaging using an infrared dammann grating," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 4, pp. 1641–1648, 2016.

[11] S. Zhang and S.-T. Yau, "High dynamic range scanning technique," *Optical Engineering*, vol. 48, no. 3, p. 033604, 2009.

[12] R. Yonesaka *et al.*, "High dynamic range digital holography and its demonstration by off-axis configuration," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 5, pp. 1658–1663, 2016.

[13] S. Feng *et al.*, "High dynamic range 3d measurements with fringe projection profilometry: a review," *Measurement Science and Technology*, vol. 29, no. 12, p. 122001, 2018.

[14] A. Łoza *et al.*, "Automatic contrast enhancement of low-light images based on local statistics of wavelet coefficients," *Digital Signal Processing*, vol. 23, no. 6, pp. 1856–1866, 2013.

[15] M. Li *et al.*, "Structure-revealing low-light image enhancement via robust retinex model," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2828–2841, 2018.

[16] K. Zhang *et al.*, "Ffdnet: Toward a fast and flexible solution for cnn-based image denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608–4622, 2018.

[17] R. Sakurai *et al.*, "Restoring aspect ratio distortion of natural images with convolutional neural network," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 563–571, 2018.

[18] V. Badrinarayanan *et al.*, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[19] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *International Conference on Learning Representations*, 2016.

[20] L.-C. Chen *et al.*, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2017.

[21] Q. Chen *et al.*, "Fast image processing with fully-convolutional networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2497–2506.

[22] L. C. Chen *et al.*, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[23] K. Men *et al.*, "Cascaded atrous convolution and spatial pyramid pooling for more accurate tumor target segmentation for rectal cancer radiotherapy," *Physics in Medicine & Biology*, vol. 63, no. 18, p. 185016, 2018.

[24] M. A. Martínez *et al.*, "Adaptive exposure estimation for high dynamic range imaging applied to natural scenes and daylight skies," *Applied Optics*, vol. 54, no. 4, pp. B241–B250, 2015.

[25] Z. Zhang *et al.*, "Active exposure control for robust visual odometry in hdr environments," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 3894–3901.

[26] I. Shim *et al.*, "Gradient-based camera exposure control for outdoor mobile platforms," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 6, pp. 1569–1583, 2018.

[27] J. Kim *et al.*, "Exposure control using bayesian optimization based on entropy weighted image gradient," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 857–864.

[28] X.-J. Mao *et al.*, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, ser. NIPS16, 2016, p. 28102818.

[29] Y. Deng *et al.*, "A sequence-to-sequence deep learning architecture based on bidirectional gru for type recognition and time location of combined power quality disturbance," *IEEE Transactions on Industrial Informatics*, 2019.

[30] E. Fan, "Extended tanh-function method and its applications to nonlinear equations," *Physics Letters A*, vol. 277, no. 4-5, pp. 212–218, 2000.

[31] L. Ekstrand and S. Zhang, "Autoexposure for three-dimensional shape measurement using a digital-light-processing projector," *Optical Engineering*, vol. 50, no. 12, p. 123603, 2011.

[32] Z. Song *et al.*, "A high dynamic range structured light means for the 3d measurement of specular surface," *Optics and Lasers in Engineering*, vol. 95, pp. 8–16, 2017.
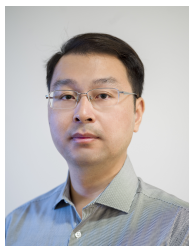
Xingjian Liu received the B.E. degree in Electronic Information Engineering from Huazhong University of Science and Technology, Wuhan, China, in 2013, and the Ph.D. degree in Materials Processing Engineering from the same university in 2018. He is currently working as a postdoctoral fellow at the Department of Mechanical and Industrial Engineering and the Robotics Institute, University of Toronto, Toronto, Canada. His interests include computer vision, 3D sensing and metrology, and robotics.

Wenyuan Chen received the B.E. degrees in computer science from the Huazhong University of science and technology, Wuhan, China, in 2018. He is currently working toward the M.A.Sc degree in Electrical and Computer Engineering with the University of Toronto, Toronto, ON, Cannada.

His research interests include computer vision, deep learning and robotics.

Harikrishnan Madhsudanan received the B.Tech., degree in Mechatronics Engineering from SRM University, Chennai, India, in 2016. He worked as an Engineer with Fiat Chrysler Automobiles R&D, Chennai, India from 2016-2018. He is currently working toward the M.A.Sc degree in Mechanical and Industrial Engineering with the University of Toronto, Toronto, ON, Canada.

His research interests include robotics, kinematics, computer vision and automation..

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2020.2991458, IEEE Transactions on Industrial Informatics

10

Ji Ge received the B.S. and M.S. degrees in mechatronics engineering from the Jiangxi University of Science and Technology, Ganzhou, China, in 2002 and 2005, respectively, and the Ph.D. degree in pattern recognition and intelligent system from Hunan University, Changsha, China, in 2012. He is currently a Post-Doctoral Fellow with the Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada, and founder of Soochow InnoVision Technology Co., Ltd. His current research interests include machine learning and computer vision.

Changhai Ru is a Professor at Soochow University. He received the Ph.D. degree from Harbin Institute of Technology in 2005. He became an associate professor in 2005 and a full professor in 2007. His research specializes in developing micro-nano technologies and instruments for manipulating and drug spraying.

Yu Sun is a Professor at the University of Toronto. He is a Tier I Canada Research Chair and Director of the Robotics Institute. His lab specializes in developing innovative technologies and instruments for manipulating and characterizing cells, molecules, and nanomaterials. He was elected Fellow of ASME (American Society of Mechanical Engineers), IEEE (Institute of Electrical and Electronics Engineers), AAAS (American Association for the Advancement of Science), NAI (National Academy of Inventors), CAE (Canadian Academy of Engineering), and RSC (Royal Society of Canada) for his work on micro-nano devices and robotic systems.