IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 11, 2018, accepted October 5, 2018, date of publication October 11, 2018, date of current version November 8, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2875496

Three-Dimensional Hand Reconstruction by Single-Shot Structured Light Line Pattern

ZHENZHOU WANG, (Member, IEEE), AND CUNSHAN ZHANG
College of Electrical and Electronic Engineering, Shandong University of Technology, Zibo 255049, China
Corresponding author: Zhenzhou Wang (zzwangsia@yahoo.com)

ABSTRACT 3-D hand reconstruction is required in many applications, e.g., hand gesture controls, hand modeling, biomechanical analysis, humanoid robot hand design and analysis, hand surgery, and so on. In most situations, it is impossible or difficult for one to keep his hand static. Thus, real-time single-shot 3-D hand reconstruction is usually required in most applications. As a result, the popular phase shift-based structured light methods could not meet the requirement, because they need at least three images to reconstruct the hand. In this paper, we propose a single-shot 3-D hand reconstruction approach based on the structured light line pattern. The proposed approach calculates a line coordinate shift map from the distorted line pattern in a single image and then reconstructs the 3-D hand with the line coordinate shift map and the calibration parameters. To calculate the line coordinate shift map accurately, we propose a specific line clustering method based on the features of the hand. We also propose a slope-based compensation method to correct the distorted line coordinate shift map. Experimental results verified the effectiveness of the proposed line clustering method based on hand features, the proposed coordinate shift map correction method, and the proposed 3-D hand reconstruction approach.

INDEX TERMS Image processing, line segmentation, line clustering, three-dimensional hand reconstruction, line coordinate shift map.

I. INTRODUCTION

Three dimensional reconstruction and modeling of hand is important in many fields that include hand surgery [1], [2], biomechanical analysis [3], humanoid robot hand design [4], [5], virtual ergonomics assessment [6] and hand gesture controls [7], [8]. For the hand surgery [1], [2], computed tomography (CT) and magnetic resonance imaging (MRI) are golden standards in soft and bony tissues. However, both CT and MRI only generate separate slices instead of the overall profile or model of the three-dimensional hand. Hence, some researchers try to model the 3D hand with all the slices of CT images [9]. Unfortunately, this kind of method is extremely computational expensive while its modeling accuracy could not be guaranteed. Besides, the radiation of CT or MRI is harmful to human and scanning of CT or MRI images usually takes a great amount of time. As a result, the researchers adopted stereophotogrammetry to reconstruct the 3D hand and they believe that robust 3D hand reconstruction and visualization has the potential to precisely simulate the preoperative experience for surgery planning. The final surgery outcome could thus be improved because of a more profound understanding of the anatomical structures that are reconstructed by 3D imaging techniques. The used 3D stereophotogrammetrical camera setup includes 15 cameras and the software program. During 3D hand reconstruction, the individual is seated with the hand put in a predesigned template with the fingers in a fully spread and upright position. After the hand is in the correct posture, the template is removed by an assistant and the hand must be kept absolute still during the photographs were taken by 15 cameras from different angles [1], [2]. Although stereophotogrammetry could reconstruct the 3D hand robustly, its operation is cumbersome. It makes the patient uncomfortable during keeping the hand absolutely still for quite a while. Therefore, a real time one-shot 3D hand imaging system is more desirable and suitable in hand surgery.

Three dimensional hand reconstruction is required in hand gesture controlling which includes controlling computer games, robots, medical visualization devices and crisis management systems [8]. The requirements by hand gesture controls for the potential three-dimensional hand reconstruction method include the adjustability to the dynamically changing background, the resistibility to the interference

of different lighting conditions, real-time implementation and dependency on user and device characteristics. As can be seen, stereophotogrammetry could not be used in hand gesture controlling applications while a real time one-shot 3D hand reconstruction method is required in most situations. Hence, some researchers propose to use the stereo vision based 3D reconstruction technique to recover 3D hand postures [10], [11] while some researchers propose to use time of flight method [12] or Kinect sensor [13] to reconstruct the 3D hand. Although these methods meet the requirement of real time and one shot, their accuracies are not adequate. A similar method that was proposed in [14] used four images simultaneously captured by four cameras to reconstruct the 3D hand and the authors have considered the large reconstruction error as the major problem. Besides the vision-based interaction techniques, there is also device-based technique that uses glove-based devices to directly measure joint angles and spatial positions of the hand [15]. Unfortunately, the glove-based devices are rather imprecise, very cumbersome, preventing the user from executing natural movements and interacting with the computer intuitively and efficiently.

There is also a branch of research work that is conducted to reconstruct and model the 3D hand from one 2D image or one video [16]–[22]. In [16], the 3D hand model is fitted with the 2D hand segmented from the 2D image. In [17], the hand posture is computed by color segmentation and finger counting in the polar coordinates. In [18], the hand is first segmented and estimated from the 2D image and the corresponding hand model is found from a large hand model database by least distance criterion. In [19] and [20], the hand is segmented and tracked in the video and the 3D hand model is also estimated from the 2D tracked result. In [21], the 2D hand is segmented and searched by nearest neighbor method in 10000 3D hand model entries to find the closest 3D model. In [22], 2D information only is used for hand segmentation and hand posture recognition. The authors claimed that 2D information is insufficient in many conditions and they suggested that a laser scanner or structured light technique should be adopted to acquire the 3D information for more robust hand posture recognition. There are many real time structured light methods [23], [26], and some of them could reconstruct hand posture or gesture satisfactorily [24]. Especially, many methods are based on the articulated motion of the hand to reconstruct the 3D hand [27]–[29]. Thus, the required structured light technique should be one shot. In addition, the required structured light technique should be resistant to the inference of different light conditions since the environments are not controllable in many applications.

In this paper, we propose a one-shot 3D hand reconstruction approach that could be used in the applications described above. The proposed approach makes use of the structured light line pattern that is generated by software beforehand [30]. The bright lines in the line pattern are designed as parallel straight lines with the same intervals and become distorted by the profile of the hand after they are projected onto the surface of the hand. The distortion of one point on the straight line could also be expressed as the coordinate shift of that point in the column direction (the direction of x coordinate in this study). The 3D profile of the hand is calculated from the distortions of all the straight lines, i.e. the coordinate shifts of all the points on the straight lines in the column direction. The distorted lines are first segmented and then clustered with specifically designed clustering method based on the features of human hand. Each clustered line is fitted with a spline function and the interval between the adjacent sampling points on the fitted spline is the same as the interval between adjacent pixels in the row direction (the direction of y coordinate). The x coordinates of the points on the fitted splines are used directly to form a low-resolution line coordinate shift matrix. The low-resolution line coordinate shift matrix is then transformed to the line coordinate shift map by spline interpolation. The spline interpolation causes distortions at the edge parts of the line coordinate shift map. Therefore, we propose a slope compensation method to correct the distorted map. Finally, the 3D hand is reconstructed with the corrected line coordinate shift map and the calibration parameters. The robustness of the 3D hand reconstruction approach is verified by experimental results. Its robustness is mainly attributed to the specific clustering method for hand reconstruction and the line coordinate shift map correction method proposed in this paper.

This paper is organized as follows. The proposed single shot 3D hand reconstruction approach described in Section 2. Specially, a line clustering method based on hand features is presented and a line coordinate shift distortion correction method is proposed. Experimental results and discussions are given in Section 3. Section 4 concludes the paper.

II. THE PROPOSED APPROACH

The flowchart of the proposed approach is shown in Fig. 1 (a) and (b). In Fig. 1 (a), the flowchart demonstrates the process of calibrating the 3D imaging system with the calibration grid. At first, the original image of the calibration grid is captured without structured light lines projected onto it. The original image is segmented by the threshold selection method proposed in [31]. In the segmented binary grid image, the areas of all the segmented blobs are calculated and the 24 markers are detected based on their areas. Then, the 24 markers are divided into six groups based on their horizontal coordinates. In each group, there are four markers and they are clustered from the top to the bottom. The unique identification number of the ith marker in the jth group is calculated as $i+4\times(j-1)$, $i=1,2,3,4$; $j=1,2,3,4,5,6$. Secondly, the line image of the calibration grid is captured with structured light lines projected onto it. The lines in the image are segmented by the line segmentation method proposed in [32]. The segmented lines are clustered from top to bottom. Based on the clustered lines, the line coordinate shift map is generated, which will be formulated in this paper. Combining the clustered markers and the line coordinate shift map of the grid image, the 3D imaging is calibrated and the system parameters are computed.

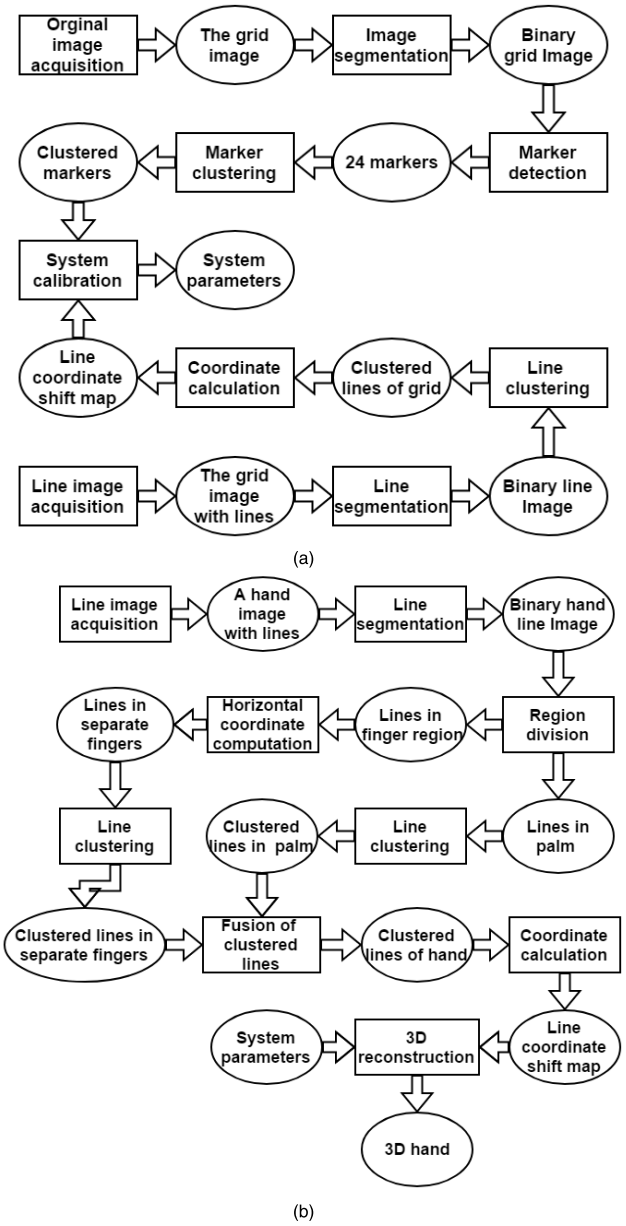


FIGURE 1. Flowchart of the proposed approach.

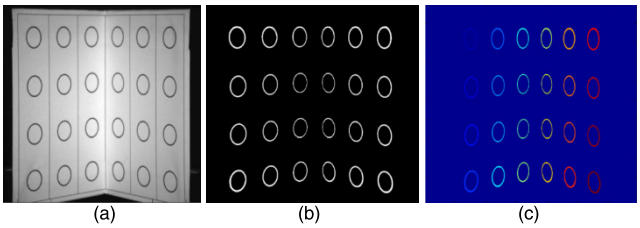


FIGURE 2. Demonstration of segmentation and clustering of the markers in the calibration grid. (a) original image of the calibration grid; (b) the segmentation result; (c) the clustering result.

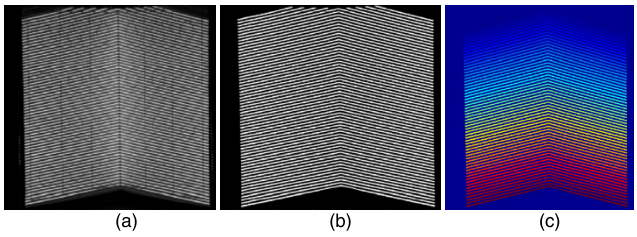


FIGURE 3. Demonstration of segmentation and clustering of the lines in the calibration grid(a) line image of the calibration grid; (b) the segmentation result; (c) the clustering result.

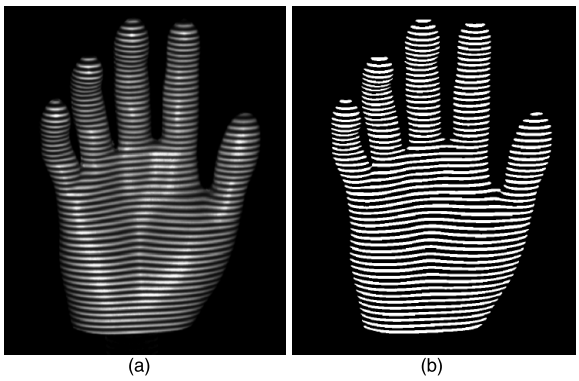


FIGURE 4. Demonstration of image segmentation (a) the captured hand image; (b) the segmentation result.

In Fig. 1 (b), the flowchart demonstrates the process of reconstructing the 3D hand with one single image. Firstly, the structured light line pattern is projected onto the hand and an image is captured. The lines in the hand image are segmented by the line segmentation method proposed in [32]. Based on the areas of the segmented lines, the lines belonging to the palm region and the lines belonging to the finger region are computed. The lines in the palm region are clustered from bottom to top. Based on the horizontal coordinates of the lines in the finger region, the lines are divided into five groups that belong to separate fingers. In each finger group, the lines are clustered independently from bottom to top. After all the lines are clustered, they are fused to form the clustered lines of the hand. Based on the clustered lines of the hand, the line coordinate shift map is generated. With the line coordinate shift map and the calibration parameters, the 3D hand is reconstructed with a single image in real time.

As can be seen, segmentation and clustering are the most critical techniques for the proposed approach. Fig. 2 shows the segmentation and clustering results of the markers in the calibration grid. The segmentation and clustering methods have been addressed in [30] and [31] and thus they are not the focus of this paper. Fig. 3 shows the segmentation and clustering results of the lines in the calibration grid. The segmentation and clustering methods have been addressed in [30] and [32]. The clustering process is simply from top to bottom based on the minimum vertical coordinate of the top line. Generation of the line coordinate shift map from the clustered lines will be formulated in this paper. Fig. 4 shows the segmentation result of the lines in the hand by the method proposed in [32]. The clustering methods for the lines in the hand are much more challenging than the method of

clustering the lines in the calibration grid. It will be proposed and described in detail in the following subsections.

A. LINE CLUSTERING FROM BOTTOM TO TOP

One important feature of the hand is that its surface is smooth, which makes the segmented lines on the palm or on the fingers continuous. Thus, we propose a bottom to top line clustering method based the labeling results of all the segmented lines.

Step 1: Suppose the resolution of the hand image is $N_X \times N_Y$. A cluster image C with the same resolution as the hand image is generated with all pixel values are assigned as 0. The segmented lines in the segmented (binary) hand image are labeled to form a labeled image L. Therefore, each line segment in the labeled image could be identified by the unique number i, $i = 1, 2, \ldots, N$ assigned to it and N is the total number of the line segments.

Step 2: Suppose all the indexes in the column direction of the hand image form the set, $X = \{1, 2, \ldots, N_X\}$ and all the indexes in the row direction form the set, $= \{1, 2, \ldots, N_Y\}$. The bottommost point (x_b, y_b) in the segmented hand image H is calculated as follows.

$$x_b = \max\left(\{x \,|\, x \in X, H\,(x, y) > 0, y \in Y\}\right) \tag{1}$$

Calculate the bottommost set $(\bar{X}, \bar{Y})$ that consists of all the positions (x, y) in the labeled image L whose pixel value equal the pixel value of bottommost point (x_b, y_b),

$$(\bar{X}, \bar{Y}) = \{(x, y) \,|\, x \in X, y \in Y, L\,(x, y) = L(x_b, y_b)\} \tag{2}$$

The corresponding positions belonging to the bottommost set $(\bar{X}, \bar{Y})$ in the binary hand image H are assigned 0 to remove the selected line.

$$H\,(x, y) = 0, \quad \forall (x, y) \in (\bar{X}, \bar{Y}) \tag{3}$$

The pixel value of the corresponding position (x, y) that belongs to the bottommost set $(\bar{X}, \bar{Y})$ in the cluster image C is assigned the number 1 to generate the first clustered line.

$$C\,(x, y) = 1, \quad \forall (x, y) \in (\bar{X}, \bar{Y}) \tag{4}$$

Step 3: The bottommost point (x_b, y_b) in the updated binary hand image H is calculated by Eq. 1 again. The bottommost set $(\bar{X}, \bar{Y})$ in the labeled image L is updated by Eq. 2 and the corresponding positions that belong to $(\bar{X}, \bar{Y})$ in the cluster image C are assigned with the number 2. The binary hand image H is then updated by Eq. 3. In the same way, the pixel value of the corresponding position (x, y) in the cluster image C is assigned the number i to generate the ith clustered line.

$$C\,(x, y) = i, \quad \forall (x, y) \in (\bar{X}, \bar{Y}) \tag{5}$$

Step 4: The binary hand image is then updated by Eq. 3. The bottommost point (x_b, y_b) in the updated binary hand image H is calculated by Eq. 1. The bottommost set $(\bar{X}, \bar{Y})$ in the labeled image L is updated by Eq. 2 and the corresponding positions that belong to $(\bar{X}, \bar{Y})$ in the cluster image C are assigned with the value calculated by adding 1 to the previous number.

$$C\,(x, y) = i + 1, \quad \forall (x, y) \in (\bar{X}, \bar{Y}) \tag{6}$$

Step 5: Repeat **Step 4** until all the lines in the binary hand image H are moved to the cluster image C with ordered numbers from the bottom to the top.

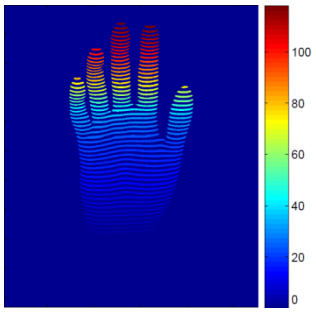


FIGURE 5. Line clustering result by the bottom to top method.

B. PROBLEM OF GENERAL LINE CLUSTERING METHODS

The general line clustering methods are proposed to meet the line clustering requirements by different applications. Therefore, the clustering rules of the general methods are fixed and they themselves could not incorporate the features of the specific applications into their clustering processes. When the situation does not follow the regularity by which the clustering method is designed, the clustering method will fail. For instance, the slope based clustering method proposed in [30] could not cluster a line broken into several line segments when the distance between two adjacent segments is too large to make the slope continuous. The challenges for the slope based method to cluster different line segments separated by the spread fingers with certain distances include: (1), it is difficult to find a starting point for the line segments separated by the fingers; (2), the slopes of these separated line segments change abruptly at the edges of the fingers. Consequently, the slope based method might not be a good option for 3D hand reconstruction.

Similarly, the bottom to top clustering method that is formulated in the above section could not cluster the broken lines correctly because they will cluster two broken parts of the same line as two different lines. During hand reconstruction, the fingers might break one line into several parts and the number of the clustered lines by bottom to top clustering method will be much higher than the practical number of the projected lines. As shown in Fig. 5, the number of the clustered lines by the bottom to top method is 118 while the practical number is 52. However, the lines on the palm parts are clustered correctly because of their continuities. In the same way, the lines on the same finger could also be clustered correctly by bottom to top method. Therefore, the bottom to top clustering method could be incorporated into the specific line clustering method designed for hand reconstruction.

C. LINE CLUSTERING BASED ON THE FEATURES OF HAND

From the above analysis, it is seen that the general clustering method alone could not cluster the lines on the hand correctly. To address the clustering problem for hand reconstruction effectively, we need the specifically designed clustering method based on the features of hand. Thus, we propose a heuristic clustering method as follows:

Step 1: Similar to the start of the bottom to top clustering method, the segmented lines in the segmented hand image are labeled to form a labeled image L. Each line segment in L is identified by the unique number $i, i = 1, 2, \ldots, N$ and N is the total number of the line segments. The coordinate set (X_i, Y_i) of each labeled line segment is calculated as:

$$(X_i, Y_i) = \{(x, y) \mid x \in X, y \in Y, L(x, y) = i\} \tag{7}$$

The size threshold to separate the lines on the fingers and the lines on the palm is calculated as:

$$\mathrm{T}_S = \frac{1}{N}\sum_{i=1}^{N} |X_i| = \frac{1}{N}\sum_{i=1}^{N} |Y_i| \tag{8}$$

$|X_i|$ or $|Y_i|$ denotes the total number of the element contained in set X_i or Y_i, i.e. the total number of pixels in the ith labeled line.

The coordinate set (X_F, Y_F) of labeled lines on the fingers is calculated as:

$$(X_F, Y_F) = \bigcup_{i=1}^{N} \{(x, y) \mid x \in X, y \in Y, |X_i| < \mathrm{T}_S\} \tag{9}$$

The coordinate set (X_P, Y_P) of labeled lines on the palm is calculated as:

$$(X_P, Y_P) = \bigcup_{i=1}^{N} \{(x, y) \mid x \in X, y \in Y, |X_i| > \mathrm{T}_S\} \tag{10}$$

A finger image F with the resolution is $N_X \times N_Y$ is generated as:

$$F(x, y) = 0, \quad \forall (x, y) \in (X, Y) \tag{11}$$

The finger image F is updated with the coordinate set (X_F, Y_F):

$$F(x, y) = 1, \quad \forall (x, y) \in (X_F, Y_F) \tag{12}$$

A palm image P with the resolution is $N_X \times N_Y$ is generated as.

$$P(x, y) = 0, \quad \forall (x, y) \in (X, Y) \tag{13}$$

The palm image P is updated with the coordinate set (X_P, Y_P):

$$P(x, y) = 1, \quad \forall (x, y) \in (X_P, Y_P) \tag{14}$$

The lines in the palm image P are clustered by the bottom to top method to form the clustered palm image P_C. Fig. 6 (a) shows the clustered result and there are 29 lines in the palm image.

Step 2: The lines in the finger image F are labeled to form a labeled image L_F. Fig. 6 (b) shows the labeling result and the total number of lines is $N_F = 89$. The coordinate set (X_i^F, Y_i^F) of the ith labeled line segment is calculated as:

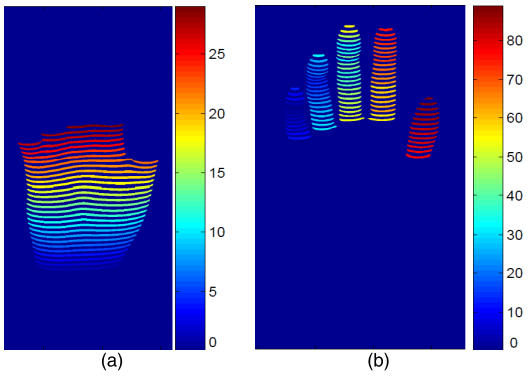


FIGURE 6. Demonstration of the heuristic clustering method based on features of hand (a) the clustered lines on the palm; (b) the labeled lines on the fingers.

$$(X_i^F, Y_i^F) = \{(x, y) \mid x \in X, y \in Y, L_F(x, y) = i\} \tag{15}$$

The average y coordinate of the ith labeled line segment in L_F is calculated as:

$$\overline{\mathrm{y}}_i = \frac{1}{N_i}\sum_{i=1}^{N_i} y_i^F, y_i^F \in Y_i^F, N_i = \left|Y_i^F\right| \tag{16}$$

$\left|Y_i^F\right|$ denotes the total number of the element contained in set Y_i^F, i.e. the total number of pixels in the ith labeled line.

The calculated N_F average y coordinates constitutes of the set $\overline{Y}_i^F = \left\{\overline{\mathrm{y}}_1, \overline{\mathrm{y}}_2, \ldots, \overline{\mathrm{y}}_i, \ldots, \overline{\mathrm{y}}_{N_F}\right\}$ for the N_F labeled lines in the image L_F. N_F equals 89 in this specific case. These average y coordinates belong to five groups: (1) little finger; (2) ring finger; (3) middle finger; (4) index finger; (5) thumb. To partition them, the differences $\mathrm{d}_F = \{d_1, d_2, \ldots d_{N_F-1}\}$ between adjacent average y coordinates are calculated:

$$d_i = \overline{\mathrm{y}}_{i+1} - \overline{\mathrm{y}}_i, \quad i = 1, 2, \ldots, N_F - 1 \tag{17}$$

The positions where four peaks occur are the partition points between the five groups. The positions $p_j, j = 1, 2, 3, 4$ of the four peaks are calculated in ascending order by the follow equations:

$$p_j = \mathrm{argmax}_{i=1,2\ldots N_{F-1}}\, d_i \tag{18}$$

$$d_{p_j} = 0 \tag{19}$$

The partition points $P_j, j = 4, 3, 2, 1$ are respectively used to separate the index finger and the thumb, to separate the middle finger and the index finger, to separate the ring finger and the middle finger, to separate the little finger and the ring finger. They are calculated in the descending order by the following equations:

$$i_j = \mathrm{argmax}_{i=1,2,3,4}\, p_i \tag{20}$$

$$P_j = p_{i_j} \tag{21}$$

$$p_{i_j} = 0 \tag{22}$$

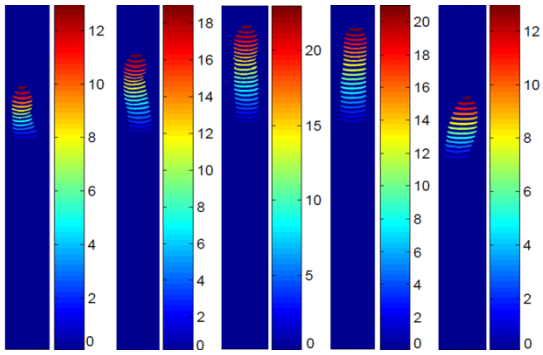


FIGURE 7. The respectively clustered lines on five fingers.

The coordinate sets $(X_{fj}, Y_{fj}), j = 1, 2, 3, 4, 5$ on the little finger, the ring finger, on the middle finger, on the index finger and on the thumb finger are calculated by the following equations:

$$(X_{fj}, Y_{fj}) = \begin{cases} \{(x, y) \mid x \in X, y \in Y, L_F(x, y) \leq P_j\}, & if\ j = 1 \\ \{(x, y) \mid x \in X, y \in Y, L_F(x, y) > P_j\}, & if\ j = 5 \\ \{(x, y) \mid x \in X, y \in Y, P_{j-1} < L_F(x, y) \leq P_j\}, & else \end{cases} \tag{23}$$

An image $F_j, j = 1, 2, 3, 4, 5$ with the resolution $N_X \times N_Y$ is generated as the little finger image, ring finger image, middle finger image, index finger image and thumb finger image respectively by the following equation.

$$F_j(x, y) = 0, \quad \forall (x, y) \in (X, Y) \tag{24}$$

Then, the image $F_j, j = 1, 2, 3, 4, 5$ is updated with the coordinate set (X_{fj}, Y_{fj}) respectively by the following equation.

$$F_j(x, y) = 1, \quad \forall (x, y) \in (X_{fj}, Y_{fj}) \tag{25}$$

The lines in these five finger images are clustered by the bottom to top method to generate five clustered finger images $F_{Cj}, j = 1, 2, 3, 4, 5$. Fig. 7 shows the clustered results for these five finger images respectively.

Step 3: After the lines in the finger part and in the palm part are clustered, they should be combined and numbered from bottom to top continuously. The numbers of the lines in the palm part will be kept while the numbers of the lines in the finger part need to be increased based on the position where they connect the palm part. To detect the position where the palm part and the fingers are connected, the lengths of the clustered lines in the clustered palm image P_C are computed. The rightmost end of the ith clustered line is calculated as:

$$M_r(i) = \max(\{y \mid P_C(x, y) = i, x \in X, y \in Y\}) \tag{26}$$

The leftmost end of the ith clustered line is calculated as:

$$M_l(i) = \min(\{y \mid P_C(x, y) = i, x \in X, y \in Y\}) \tag{27}$$

Then the lengths of the ith clustered lines are approximated with the following equation.

$$L_P(i) = M_r(\mathrm{i}) - M_l(\mathrm{i}), \quad i = 1, 2, \ldots N_P \tag{28}$$

Where N_P is the total number of the clustered lines in P_C. The differences $\mathrm{D}_P = \{D_1, D_2, \ldots D_{N_P-1}\}$ between adjacent lengths are calculated:

$$D_i = L_P(i + 1) - L_P(i), \quad i = 1, 2, \ldots, N_P - 1 \tag{29}$$

The positions where N_C peaks occur are the connecting points to connect the fingers and the palm. N_C ranges from 2 to 4 and it equals 3 for the demonstrated hand. The positions $p'_j, j = 1, \ldots, N_C$ of the N_C peaks are calculated in the ascending order by the following equations.

$$p'_j = \mathrm{argmax}_{i=1,2\ldots N_{P-1}} D_i \tag{30}$$

$$D_{p'_j} = 0 \tag{31}$$

For the case $N_C = 4$, the connecting point $P'_j, j = 4, \ldots, 1$ to connect the index finger and the palm, to connect the ring finger and the palm, to connect the little finger and the palm and to connect the thumb and the palm are calculated by the following equations:

$$i_j = \mathrm{argmax}_{i=1,2,3} p'_i \tag{32}$$

$$P'_j = p'_{i_j} \tag{33}$$

$$p'_{i_j} = 0 \tag{34}$$

The connecting point $P'_j, j = 5$ to connect the middle finger and the palm is calculated as:

$$P'_j = N_P \tag{35}$$

For the case $N_C = 3$, the connecting point $P'_j, j = 3, \ldots, 1$ to connect the ring finger and the palm, to connect the little finger and the palm and to connect the thumb and the palm are calculated by Eqs. (32)-(34). The connecting point P'_j, $j = 5$ to connect the middle finger and the palm, the connecting point $P'_j, j = 4$ to connect the index finger and the palm are calculated by Eq. (35).

For the case $N_C = 2$, the connecting point P'_j, $j = 2, \ldots, 1$ to connect the little finger and the palm and to connect the thumb and the palm are calculated by Eqs. (32)-(34). The connecting point $P'_j, j = 5$ to connect the middle finger and the palm, the connecting point $P'_j, j = 4$ to connect the index finger and the palm, the connecting point $P'_j, j = 3$ to connect the ring finger and the palm are calculated by Eq. (35).

The assigned values of the clustered lines in the image F_{Cj}, $j = 1, 2, 3, 4, 5$ are updated by the following equations.

$$(X_{Cj}, Y_{Cj}) = \left\{(x, y) \mid x \in X, y \in Y, F_{Cj}(x, y) > 0\right\} \tag{36}$$

$$F_{Cj}(x, y) = F_{Cj}(x, y) + P'_j, \quad \forall (x, y) \in (X_{Cj}, Y_{Cj}) \tag{37}$$

Then, the five updated finger images are combined to generate the clustered finger image F_C which is initialized as:

$$F_C(x, y) = 0, \quad \forall (x, y) \in (X, Y) \tag{38}$$

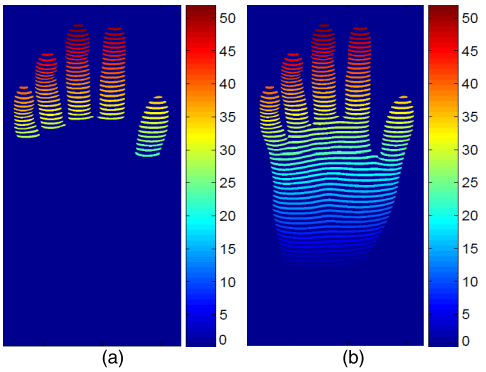


FIGURE 8. Final results of the heuristic clustering method (a) the clustered finger image; (b) the clustered hand image.

and then updated as:

$$F_C(x,y) = \begin{cases} F_{C1}(x,y), & \forall (x,y) \in (X_{C1}, Y_{C1}) \\ F_{C2}(x,y), & \forall (x,y) \in (X_{C2}, Y_{C2}) \\ F_{C3}(x,y), & \forall (x,y) \in (X_{C3}, Y_{C3}) \\ F_{C4}(x,y), & \forall (x,y) \in (X_{C4}, Y_{C4}) \\ F_{C5}(x,y), & \forall (x,y) \in (X_{C5}, Y_{C5}) \end{cases} \quad (39)$$

Fig. 8 (a) shows the clustered finger image F_C. F_C is combined with P_C to generate the final clustered hand image H_C.

$$H_C(x,y) = \begin{cases} P_C(x,y), & \forall (x,y) \in (X_P, Y_P) \\ F_C(x,y), & \forall (x,y) \in (X_F, Y_F) \end{cases} \quad (40)$$

$$(X_F, Y_F) = \bigcup_{i=1}^{5} (X_{Ci}, Y_{Ci}) \quad (41)$$

Fig. 8 (b) shows the final clustered hand image H_C.

D. LINE COORDINATE SHIFT MAP CALCULATION AND CORRECTION

The skeletons of all the clustered lines in the final clustered hand image H_C are computed as follows:

Step 1: A line coordinate shift matrix M with the resolution $N \times N_Y$ is generated as:

$$M(x,y) = 0, \quad \forall (x,y) \in (X_N, Y) \quad (42)$$

where $X_N = \{1, 2, \ldots, N\}$.

Step 2: A line image with the resolution $N_X \times N_Y$ is generated as:

$$l(x,y) = 0, \quad \forall (x,y) \in (X, Y) \quad (43)$$

The ith clustered line in the final clustered hand image H_C is moved to the line image by the following equations.

$$(X_i^H, Y_i^H) = \{(x,y) \mid x \in X, y \in Y, H_C(x,y) = i\} \quad (44)$$

$$l(x,y) = 1, \quad \forall (x,y) \in (X_i^H, Y_i^H) \quad (45)$$

The skeleton of the ith clustered line is then calculated as:

$$S_i = \bigcup_{n=1}^{N_B} \{(l \ominus nB) - (ln \ominus B) \circ\} \quad (46)$$

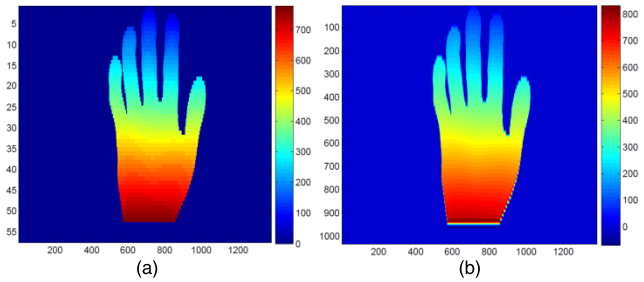


FIGURE 9. Demonstration of generating line coordinate shift map (a) the generated matrix; (b) the generated line coordinate shift map.

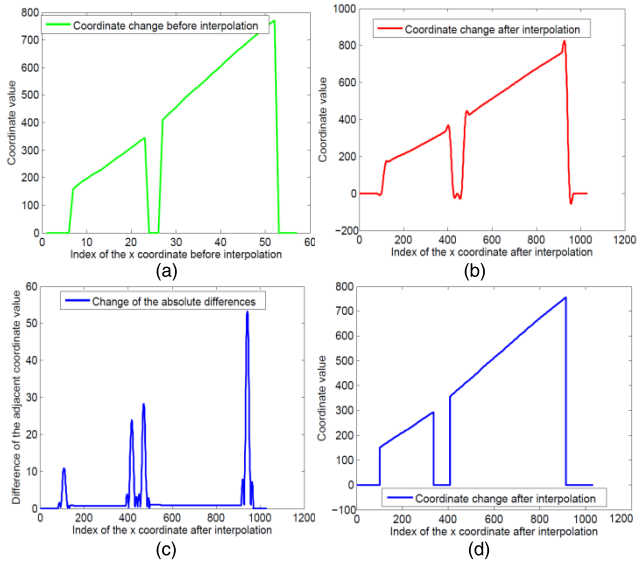


FIGURE 10. Demonstration of line coordinate shift map correction (a) the original coordinate shift; (b) the distorted coordinate shift by polynomial interpolation; (c) absolute differences of the distorted coordinate shift; (d) the corrected coordinate shift.

where B is the structuring element and N_B is the size of the structuring element. The ith row of the line coordinate shift matrix is updated as:

$$M(i,j) = S_i(j), \quad j = 1, 2, \ldots, N_Y \quad (47)$$

Step 3: Repeat **Step 2** from $i = 1$ until it reaches N.

Fig. 7 (a) shows the generated line coordinate shift matrix M. We use the spline interpolation to transform the line coordinate shift matrix from the resolution $N \times N_Y$ to the resolution $N_X \times N_Y$ to generate the line coordinate shift map M_s. Fig. 7 (b) shows the generated line coordiante shift map M_s. It is seen that the color-bar in Fig. 9 (a) ranges from 0 to less than 800 while the color-bar in Fig. 9 (b) ranges from less than 0 to more than 800, which indicates that the line coordinate shift map is distorted during the polynomial interpolation process.

To see the distortion clearly, we plot one column of the shift matrix in Fig. 10 (a) and we plot the same column of the shift map in Fig. 10 (b). As can be seen, the edge parts in the matrix change to zeros abruptly while the edge parts in the map change to zeros along a smoothing spline curve.

In addition, the minimum and maximum values of the map are distorted by the spline curves. In the coordinate shift map, the spline interpolation between zero values and large non-zero values causes the distortion. To rectify the coordinate shift distortion, a slope based correction method is proposed as follows.

Step 1: for each column of the line coordinate shift map, l_c its absolute differences are computed as demonstrated in Fig. 10 (c). At the edge of the hand, the difference yields one major peak with large magnitude and two minor peaks with small magnitudes around it. The position of the major peak is calculated as $K_i, i = 1, 2, ..N_K$ and the intervals between the major peak and its corresponding two minor peaks are computed I_i^l and I_i^r respectively. N_K denotes the total number of the major peaks.

Step 2: for the $(K_i - I_i^l + j)$th point $(K_i - I_i^l + j, l_c\left(K_i - I_i^l + j\right))$ on the selected column l_c, ten points on its left are selected to fit a straight line.

$$y_i^l = a^l x_i^l + b^l \tag{48}$$

$$\left[a^l, b^l\right]^T = \left(B_1^T B_1\right)^{-1} B_1^T E_1 \tag{49}$$

$$B_1 = \begin{bmatrix} K_i - I_i^l + j - 10 & 1 \\ K_i - I_i^l + j - 9 & 1 \\ \vdots & \vdots \\ K_i - I_i^l + j - 1 & 1 \end{bmatrix} \tag{50}$$

$$E_1 = [l_c(K_i - I_i^l + j - 10), \cdots, l_c(K_i - I_i^l + j - 1)]^T \tag{51}$$

Then the $(K_i - I_i^l + j)$th point is updated as:

$$y^l_{(K_i - I_i^l + j)} = a^l(K_i - I_i^l + j) + b^l \tag{52}$$

Step 3: repeat **Step 2** from j equals 0 until it reaches I_i^l.

Step 4: for the $(K_i + I_i^r - j)$th point $(K_i + I_i^r - j, l_c\left(K_i + I_i^r - j\right))$ on the selected column l_c, ten points on its right are selected to fit a straight line.

$$y_i^r = a^r x_i^r + b^r \tag{53}$$

$$\left[a^r, b^r\right]^T = \left(B_2^T B_2\right)^{-1} B_2^T E_2 \tag{54}$$

$$B_2 = \begin{bmatrix} K_i + I_i^r - j + 1 & 1 \\ K_i + I_i^r - j + 2 & 1 \\ \vdots & \vdots \\ K_i + I_i^r - j + 10 & 1 \end{bmatrix} \tag{55}$$

$$E_2 = \left[l_c\left(K_i + I_i^r - j + 1\right), \cdots, l_c\left(K_i + I_i^r - j + 10\right)\right]^T \tag{56}$$

Then the $(K_i + I_i^r - j)$th point is updated as:

$$y^r_{(K_i + I_i^r - j)} = a^r(K_i + I_i^r - j) + b^r \tag{57}$$

Step 5: repeat **Step 4** from j equals 0 until it reaches I_i^r.

Fig. 10 (d) shows the corrected coordinate shift. As can be seen, the shape of the corrected phase is very similar to the shape of the original coordinate shift as shown in Fig. 10 (a),

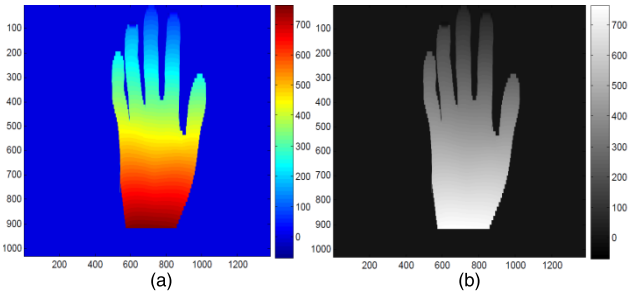


FIGURE 11. Demonstration of the generated line coordinate shift map (a) color map; (b) gray map.

which indicates that the proposed method is effective. Fig. 11 (a) and (b) show the corrected line coordinate shift map and the color-bar of the corrected map ranges the same as the coordinate shift matrix as shown in Fig. 9 (a). With the line coordinate shift map and the calibration parameters, the 3D hand is reconstructed as described in [30].

III. RESULTS AND DISCUSSION

We calculate the reconstruction accuracy of the established system with 24 points on the calibration grid [30] and the average mean squared error (MSE) is only 0.38 mm. Since the size of the calibration grid is about 200 mm, the percentage error is about 0.2%, which is similar to that of traditional phase shift profilometry [33] and significantly more accurate than other one-shot projection based state-of-the-art 3D imaging methods [34], [35]. The proposed approach is implemented with VC++ in an i7-6700, 3.4 GHz computer and the processing time of reconstructing a 3D 1280×1024 hand is 156.3 milliseconds, which is promising for the real time applications.

Fig. 12 shows the results of reconstructing the hand at different angles. The images in the first column are the two-dimensional images acquired by the camera. The images in the second column are the 3D reconstruction results by the proposed 3D hand reconstruction approach without coordinate shift correction. The images in the third column are the 3D reconstruction results by the proposed approach with coordinate shift correction. As can be seen, the results in the third column look good, which verifies the effectiveness of both the proposed one-shot 3D hand reconstruction approach and the proposed coordinate shift correction method. For comparison purpose, we show the results of state-of-the-art single-shot structured light methods [34], [35] in Fig. 13. As can be seen, the 3D hand reconstructed by the proposed approach appears to be better than those by [34] and [35].

The hand with the fingers spread out is more challenging for the 3D reconstruction method than the hand in other poses. By comparison, we see that the proposed single shot 3d hand reconstruction approach is more robust than state-of-the-art single shot structured light methods [34], [35]. The proposed approach is promising for hand surgery [1], [2], hand gesture controls [7], [8] and other applications [3], [6].

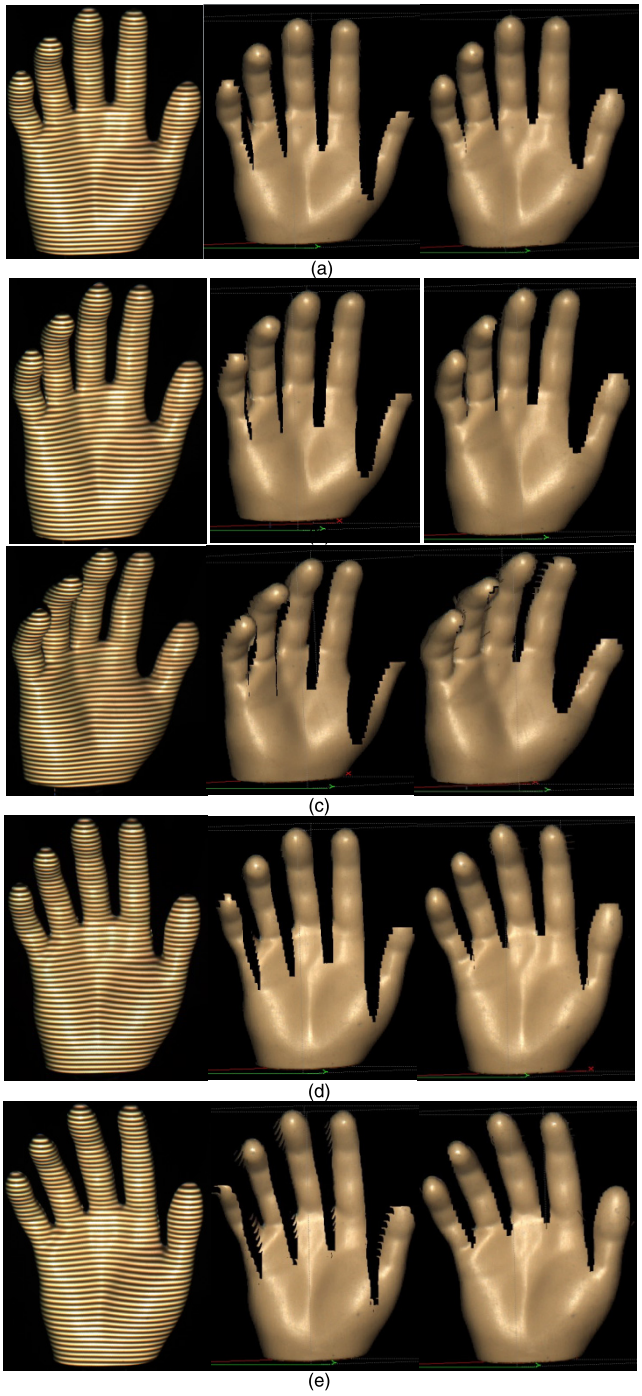


FIGURE 12. Results of reconstructing the hand by the proposed one-shot approach at different angles; (a) results at angle 1; (b) results at angle 2; (c) results at angle 3; (d) results at angle 4; (e) results at angle 5. (From left to the right, the images are the captured two-dimensional images, the reconstructed 3D hand without coordinate shift correction and the reconstructed 3D hand with coordinate shift correction respectively).

In summary, the major contributions of this paper include:

(1), a one-shot 3D hand reconstruction method based on image processing technology and the features of hand is proposed and its effectiveness is validated in this paper.

(2), a bottom to top clustering algorithm is formulated to cluster the continuous lines robustly.

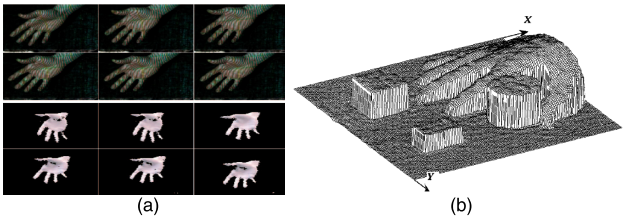


FIGURE 13. Reconstructed hands by state-of-the-art single shot structured light methods (a) results reconstructed in [34]; (b) results reconstructed in [35].

(3), a heuristic line clustering method based on the features of hand is proposed to cluster the lines on the spread-out hand specifically.

(4), a slope based line coordinate shift distortion correction method is proposed and its effectiveness is verified in this paper.

(5), a line coordinate shift map computation method based on the clustered lines in a single image is formulated in this paper.

IV. CONCLUSION AND FUTURE WORK

In this paper, a single shot 3D hand reconstruction approach is proposed and its effectiveness is verified by experimental results. The 3D coordinates of the hand are computed based on the 2D line coordinate shift map calculated from one single image instead of more than three images by phase shift methods. To compute the coordinate shift map robustly, the gradients of lines are computed, segmented and then clustered based on hand features. Because the line gradient is used for segmentation, the proposed approach is more resistant to light inferences than other structured light methods. Thus, it is promising to meet the requirements of outdoor applications.

Single shot and robust reconstruction of the 3D hand is demanded by a variety of applications and the poses of the hand might vary significantly in different applications. Thus, we will improve the proposed approach to be able to reconstruct the 3D hand with different poses robustly in the near future.

REFERENCES

[1] I. A. Hoevenaren, T. J. J. Maal, E. Krikken, A. F. J. de Haan, S. J. Bergé, and D. J. O. Ulrich, "Development of a three-dimensional hand model using 3D stereophotogrammetry: Evaluation of landmark reproducibility," *J. Plastic, Reconstruct. Aesthetic Surg.*, vol. 68, no. 5, pp. 709–716, 2015.

[2] I. A. Hoevenaren, J. Meulstee, E. Krikken, S. J. Berge, D. J. O. Ulrich, and T. J. Maal, "Development of a three-dimensional hand model using three-dimensional stereophotogrammetry: Assessment of image reproducibility," *plos one*, vol. 10, no. 9, p. e0136710, 2015.

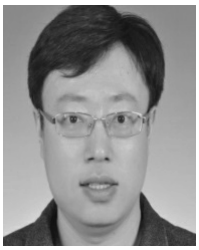
[3] K. N. An, E. Y. Chao, W. P. Cooney, III, and R. L. Linscheid, "Normative model of human hand for biomechanical analysis," *J. Biomech.*, vol. 12, no. 10, pp. 775–788, 1979.

[4] J. Hirt, K. Berns, and K. Mianowski, "Designing arms and hands for the humanoid robot ROMAN," *Adv. Mater. Res.*, vols. 463–464, pp. 1233–1237, Feb. 2012.

[5] I. Virgala, M. Kelemen, M. Varga, and P. Kuryło, "Analyzing, modeling and simulation of humanoid robot hand motion," *Procedia Eng.*, vol. 96, pp. 489–499, Jan. 2014.

[6] Y. Endo *et al.*, "Optimization-based grasp posture generation method of digital hand for virtual ergonomics assessment," *SAE Int. J. Pas-Senger Cars-Electron. Elect. Syst.*, vol. 1, no. 1, pp. 590–598, 2009.

[7] J. P. Wachs, M. Kölsch, H. Stern, and Y. Edan, "Vision-based hand-gesture applications," *Commun. ACM*, vol. 54, no. 2, pp. 60–71, 2011.
[8] Y. Wu and T. S. Huang, "Human hand modeling, analysis, and animation in the context of HCI," in *Proc. Int. Conf. Image Process.*, 1999, pp. 6–10.
[9] T. Kurihara and N. Miyata, "Modeling deformable human hands from medical images," in *Proc. ACM SIGGRAPH Symp. Comput. Animation (SCA)*, Grenoble, France, 2004, pp. 357–366.
[10] A. A. Argyros and M. I. A. Lourakis, "Binocular hand tracking and reconstruction based on 2D shape matching," in *Proc. Int. Conf. Pattern Recognit.*, vol. 1, 2006, pp. 207–210.
[11] S. Grange, T. Fong, and C. Baur, "M/ORIS: A medical/operating room interaction system," in *Proc. ACM Int. Conf. Multimodal Interfaces*, 2004, pp. 159–166.
[12] G. Hackenberg, R. McCall, and W. Broll, "Lightweight palm and finger tracking for real-time 3D gesture control," in *Proc. IEEE Virtual Reality Conf. (VR)*, Singapore, Mar. 2011, pp. 19–26.
[13] K. Kim, J. Kim, J. Choi, J. Kim, and S. Lee, "Depth camera-based 3D hand gesture controls with immersive tactile feedback for natural mid-air gesture interactions," *Sensors*, vol. 15, no. 1, pp. 1022–1046, 2015.
[14] E. Ueda, Y. Matsumoto, M. Imai, and T. Ogasawara, "A hand-pose estimation for vision-based human interfaces," *IEEE Trans. Ind. Electron.*, vol. 50, no. 4, pp. 676–684, Aug. 2003.
[15] R. Rosales, V. Athitsos, L. Sigal, and S. Scarloff, "3D hand pose reconstruction using specialized mappings," in *Proc. 8th Int. Conf. Comput. Vis.*, 2001, pp. 378–385.
[16] J. J. Kuch and T. S. Huang, "Vision based hand modeling and tracking for virtual teleconferencing and tele-collaboration," in *Proc. 5th Int. Conf. Comput. Vis.*, 1995, pp. 666–671.
[17] X. Yin and X. Zhu, "Hand posture recognition in gesture-based human-robot interaction," in *Proc. IEEE Conf. Ind. Electron. Appl.*, May 2006, pp. 1–6.
[18] V. Athitsos and S. Sclaroff, "Estimating 3D hand pose from a cluttered image," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2003, pp. 432–439.
[19] B. Stenger, P. R. S. Mendoncca, and R. Cipolla, "Model-based 3D tracking of an articulated hand," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Kauai, HI, USA, vol. 2, Dec. 2001, pp. 310–315.
[20] B. Stenger, A. Thayananthan, P. H. S. Torr, and R. Cipolla, "Model-based hand tracking using a hierarchical Bayesian filter," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1372–1384, Sep. 2006.
[21] J. Romero, H. Kjellström, and D. Kragic, "Monocular real-time 3D articulated hand pose estimation," in *Proc. IEEE-RAS Int. Conf. Humanoid Robots*, Paris, France, Dec. 2009, pp. 87–92.
[22] S. Akyol, U. Canzler, K. Bengler, and W. Hahn, "Gesture control for use in automobiles," in *Proc. MVA*, Tokyo, Japan, 2000, pp. 349–352.
[23] L. Song, S. Tang, and Z. Song, "A robust structured light pattern decoding method for single-shot 3D reconstruction," in *Proc. IEEE Int. Conf. Real-Time Comput. Robot. (RCAR)*, Okinawa, Japan, Jul. 2017, pp. 668–672.
[24] B. Li, Y. An, D. Cappelleri, J. Xu, and S. Zhang, "High-accuracy, high-speed 3D structured light imaging techniques and potential applications to intelligent robotics," *Int. J. Intell. Robot. Appl.*, vol. 1, no. 1, pp. 86–103, 2017.
[25] J. Yang and H. Chen, "The 3D reconstruction of face model with active structured light and stereo vision fusion," in *Proc. 3rd IEEE Int. Conf. Comput. Commun. (ICCC)*, Chengdu, China, Dec. 2017, pp. 1902–1906.
[26] P. Bourke, "Automatic 3D reconstruction: An exploration of the state of the art," *GSTF J. Comput.*, vol. 2, no. 3, pp. 71–74, 2018.
[27] Y. Wu and T. S. Huang, "Human hand modeling, analysis and animation in the context of HCI," in *Proc. Int. Conf. Image Process.*, 1999, pp. 6–10.
[28] Y. Endo, M. Tada, and M. Mochimaru, "Reconstructing individual hand models from motion capture data," *J. Comput. Des. Eng.*, vol. 1, no. 1, pp. 1–12, 2014.
[29] Y. Yasumuro, Q. Chen, and K. Chihara, "Three-dimensional modeling of the human hand with motion constraints," *Image Vis. Comput.*, vol. 17, no. 2, pp. 149–156, 1999.
[30] Z. Wang and Y. Yang, "Single-shot three-dimensional reconstruction based on structured light line pattern," *Opt. Lasers Eng.*, vol. 106, pp. 10–16, Jul. 2018.
[31] Z. Wang, J. Xiong, Y. Yang, and H. Li, "A flexible and robust threshold selection method," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 9, pp. 2220–2232, Sep. 2018.
[32] Z. Z. Wang, "Unsupervised recognition and characterization of the reflected laser lines for robotic gas metal arc welding," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 1866–1876, Aug. 2017.
[33] Z. Z. Wang, "Robust measurement of the diffuse surface by phase shift profilometry," *J. Opt.*, vol. 16, no. 10, p. 105407, 2014.
[34] T. P. Koninckx and L. Van Gool, "Real-time range acquisition by adaptive structured light," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 432–445, Mar. 2006.
[35] M. Takeda, Q. Gu, M. Kinoshita, H. Takai, and Y. Takahashi, "Frequency-multiplex Fourier-transform profilometry: A single-shot three-dimensional shape measurement of objects with large height discontinuities and/or surface isolations," *Appl. Opt.*, vol. 36, no. 22, pp. 5347–5354, 1997.



ZHENZHOU WANG (M'15) received the bachelor's and master's degrees from the Department of Electronics and Information Engineering, Tianjin University, China, in 2000 and 2003, respectively, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Kentucky, in 2007.

From 2007 to 2012, he was a Researcher and a Post-Doctoral Scholar with the University of Kentucky. He was selected in the "Hundred Talents Plan, A-Class" of the Chinese Academy of Sciences in 2013. He was a Research Fellow/a Professor with the Shenyang Institute of Automation till 2018. He is currently a Full Professor and the Academic Leader with the College of Electrical and Electronic Engineering, Shandong University of Technology. He has published more than 30 SCI journal papers as the first and corresponding author. He also serves as the Panelist for the National Natural Science Foundation of China.

His research interests include artificial intelligence, signal/image processing, machine/computer vision, structured light 3-D surface imaging, monitoring, and robotics technology.

CUNSHAN ZHANG was born in Shandong, China, in 1965. He received the B.S. degree in electrical engineering from Shandong Polytechnic University in 1986, the M.S. degree in power electronics technology from the Harbin Institute of Technology in 1991, and the Ph.D. degree in power system and its automation from Beijing Jiaotong University in 2006.

Since 2001, he has been a Full Professor with the Shandong University of Technology. He is currently the Dean of the College of Electrical and Electronic Engineering. He is also the Deputy Director General of the Shandong Automation Society and the Director of the Provincial Electrical Engineering Society.

Dr. Zhang has received numerous awards, including the Award of Shandong Science and Technology Progress, the Award of Science and Technology Progress of the Ministry of Education, the Shandong Provincial Education Commission Award in Science and Technology Progress, and the First Prize of Teaching Achievement in Shandong.