IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 17, 2018, accepted November 7, 2018, date of publication November 12, 2018,
date of current version December 18, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2880494

Reformed Residual Network With Sparse Feedbacks for 3D Reconstruction From a Single Image

YUJUAN SUN[1], MUWEI JIAN[2,3], AND XIAOFENG ZHANG[1]
[1]Department of Information and Electrical Engineering, Ludong University, Yantai 264025, China
[2]School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan 250014, China
[3]School of Information Science and Engineering, Linyi University, Linyi 276000, China

Corresponding author: Yujuan Sun (syj_anne@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61602229, Grant 61601427, and Grant 61873117, in part by the Natural Science Foundation of Shandong under Grant ZR2016FM13, Grant ZR2016FM40, and Grant ZR2016FM21, in part by the Shandong Provincial Key Research and Development Program of China under Grant 2017CXGC0701 and Grant 2017CXGC1504, and in part by the Fostering Project of Dominant Discipline and Talent Team of Shandong Province Higher Education Institutions.

ABSTRACT Deep neural networks are difficult to train due to the large number of unknown parameters. To increase the trainable performance, we present a novel network model with moderate depth for three-dimensional reconstruction. The proposed network model, called SFResNet, only has eight layers, and sparse feedbacks were added in the middle and last layers, which is mainly used to add the constraints and improve the stability of the network model. In addition, a joint strategy is proposed to reduce the artificial Mosaic trace at the seam of the patches; hence, SFResNet can also evaluate an input image of any size. Visually pleasing output results can be produced with a reconstructed shape and normal surface. The experimental results show the effectiveness of the proposed method.

INDEX TERMS 3D reconstruction, surface normal, residual network, sparse feedback.

I. INTRODUCTION

In recent years, many difficult problems in computer vision have been solved by using the deep learning method, and the accuracy and robustness of these methods have been considerably improved. This is especially true in the fields of image classification, object detection and identification.

Until now, many types of neural network structures have been proposed, such as LeNet 5 [3], AlexNet 8 [4], VGG [6], GoogleNet [5], ResNet [7], and GAN [8]. These networks have been gradually deepened, and the training data set is also becoming larger. Although the technology of deep learning is especially effective for issues of classification, there are still many problems that cannot be solved with 3D reconstruction. The main reason is due to two aspects: first, collecting a large number of training pairs for an image and its corresponding 3D shape is not easy; second, CNN is good at classification and probability estimation; but, the 3D shape is a continuously changing surface, and it is difficult to directly fit these functions. Therefore, when using a deep neural network to solve these problems, the training set is not usually large enough to train the deep neural network model.

In this study, we improve the network structure of ResNet [7] and propose a sparse feedback residual network called SFResNet, which includes 8 layers with sparse ''shortcut connections'' (in this article, we call the ''shortcut connections'' forward feedback). Figure 1 shows the structure of SFResNet. SFResNet only includes two forward feedbacks, which are derived from the input layer. One feedback is connected to the middle layer, and the other is connected to the last layer. The two feedbacks can provide an effective constraint for the loss function and help to train reasonable network parameters. Experimental results show that the feedbacks in SFResNet can accelerate the convergence speed of the loss function and improve the model's robustness.

II. RELATED WORKS

Surface height map estimation is an important task in 3D reconstruction. This task differs from the general scene depth estimation, and it contains more details and high-frequency characteristics. The existing methods, such as radar or others equipment methods, can only reconstruct a

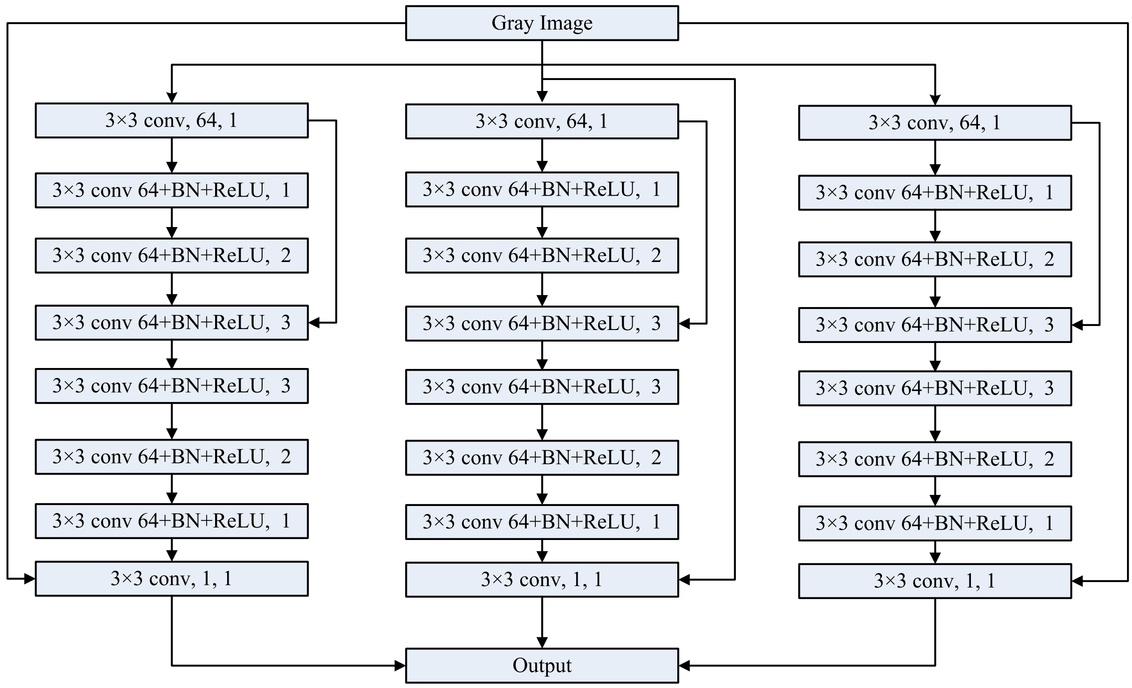

FIGURE 1. The Architecture of SFResNet.

large scale scene depth but will not work in a 3D reconstruction of small scale objects. Of course, there are many methods that have been used in 3D surface reconstruction, such as photometric stereo [12], laser triangulation [27] and shape from X [1]. These methods always establish a prior-based model for the optical imaging. Moreover, most of the prior-based methods built the objective function by simplifying the mechanism of object imaging and have not considered the effect of noise or other factors. Hence, they only can restore a height map from clean and multiple images. The surface height map estimation from one single image is still a challenging issue in computer vision. In addition, the prior-based methods involve a complex optimization problem, and most of the prior-based methods can scarcely achieve a high performance without sacrificing computational efficiency. Furthermore, the prior-based models in general are non-convex and involve several manually chosen parameters [2].

In recent years, the deep convolution neural network (CNN) [10], [11] has undergone a series of breakthroughs in many applications of computer vision [12], [13], such as image classification, recognition and target detection. The features of CNN are mainly obtained by increasing the depth of the network model; then, the lower, middle, and advanced features can be gradually obtained. In general, the advanced features will be used to connect with one or several fully connected layers. The reason for obtaining the advanced achievement in computer vision is mainly because many rich characteristics of different levels can be extracted by training the deep neural network.

Liu *et al.* proposed a method based on deep CNN [21] and CRF to estimate the depth from a single input image [25], [28]. David *et al.* predicted depth, the surface normal and semantic labels by using a common multi-scale convolutional architecture [29]. These methods are mainly used to estimate the large scale scene depth, such as for streets or natural scenes. In ICCV2017, George Trigeorgis pursued a data-driven approach to estimate the surface normal from a single intensity image [26], focusing in particular on human faces, and he trained a fully convolutional network to accurately recover facial normals. Another study [25] presented a novel method based on CNN to directly estimate the height map from a single texture image. The methods from these two references are different from the general scene depth estimation, and their estimated height maps contain more high-frequency characteristics or details.

Recent studies also revealed many visual tasks [14], [16], [17], especially low-level vision problems, which have considerably benefited from a very deep network. There are several studies [18]–[20] that have performed denoising problems using deep neural networks. Reference [18] found that a convolutional network offered similar performance using the blind denoising setting compared with other techniques using the non-blind setting. However, training a convolutional network architecture requires substantial computation and many thousands of updates to converge.

In a deep network structure, can all these extracted features be used in the best way? There may be many useless layers or useless parameters and some high-level features that may be actually useless for low-level applications of image processing. Reference [22] claimed that training a multi-layer perceptron (MLP) with many hidden layers can lead to problems, such as vanishing gradients and over-fitting,

and [22] also found that back propagation will work well and concluded that deep learning techniques are not necessary. Therefore, a moderate depth neural network is proposed in this paper, which is an 8-layered deep residual network with sparse feedback loops. Details of the proposed network will be introduced as following.

III. DEEP RESIDUAL NETWORK WITH SPARSE FEEDBACKS

In the design and application of a neural network, researchers are only required to focus on the input and output, the number of hidden layers, and the initial parameters. As the network depth gradually increases, the parameters for the neural network are also difficult to tune. There are also no relevant theories that have been presented on how to tune these parameters. Moreover, updating neuron parameters depends on the gradient: the more far away from the output layer, the more difficult it is to update the neuron parameters. It will be invariant or will dramatically change, which is called gradient disappearance or the gradient explosion problem. Although a dropout strategy or batch normalization has been adopted to reduce explosion problems, it still often occurs, and with more layers for a neural network, the gradient disappearance or explosion problem is more obvious.

This question reminds us of the amplifier cascade problem in electronics. When connecting a circuit, the output signal is usually unstable. A single negative feedback or inter-stage negative feedback will generally be added to stabilize the output signal. We found that this idea has been used in [7]. The residual learning framework (ResNet) in [7] is very similar to the negative feedback in the amplifier cascade and considerably improves the performance of the convolution neural network. Moreover, it won the first place for ILSVRC & COCO classification, ImageNet detection, ImageNet positioning, COCO detection and COCO segmentation in 2015.

However, the results of ResNet are not better than that of a convolution neural network without feedback for the problem of removing noise. For example, DnCNN [9] is good at removing Gaussian noise. Therefore, we combine the network structure of ResNet with DnCNN, and propose a deep residual network with sparse feedback loops for 3D reconstruction, which is called SFResNet. The proposed network structure is shown in Fig. 1.

The main function of the structure in Fig. 1 is to obtain the object's surface normal from its input image. Since the surface normal contains three values, the architecture shown in Fig. 1 contains three parallel branches, and each branch estimates one parameter of the surface normal. The training process can be modeled as the loss function to learn the trainable parameters, Ξ, of SFResNet as follows:

$$f(\Xi) = \frac{1}{N}\sum_{i=1}^{N} \|R(n_i, x_i, \Xi) - x_i\|^2 \quad (1)$$

where $\{(x_i,n_i)\}$ represents N image-normal training (patch) pairs; SFResNet aims to learn a mapping function $R(n_i, x_i) = y_i$, to predict the surface normal of the input image. R also represents the residual network structure, whose parameters, Ξ, must be trained. Next, we introduce the details of the proposed network structure.

A. NETWORK STRUCTURE

Inspired by the residual learning structure, we propose a deep residual network with sparse feedbacks for 3D reconstruction, and the structure of the SFResNet is shown in Fig. 1. Figure 1 contains three parallel branches, and each branch has the same network structure; however, the function of every branch is different, which is responsible for estimating a parameter of the surface normal. Each branch in Fig. 1 consists of eight layers. ''Convolution'' block is in the first layer. This layer has no ''Batch Normal'' or ''ReLU'', i.e., the information produced by this layer is the original information after filtering the input image. Then, it is used to estimate the residual information by feeding back to the middle and to the last layer. Six ''Convolution+Batch Normalization+ReLU'' blocks are in the middle layers. The numbers behind each middle layer are the dilation factors, which are set to 1, 2, 3, 3, 2 and 1, respectively. By using increasing dilated factors, the first half layers can learn the residual information using an enlarged receptive field, and the latter half layers can refine the residual information using the decreasing dilation factors. To ensure that the estimated information does not considerably deviate, two forward feedbacks from the first layer were added. The first is connected to the middle of the dilation convolution. The second is connected to the last layer. The main task of SFResNet is to estimate the residual information between the input image and the output surface normal.

B. IMPLEMENTATION

Due to the limitation of training pairs, we must reduce the size and parameters of the neural network. We cut the input training image into small patches to increase the number of training pairs and added the patches into the network for training. Of note, the dilated convolution with dilatation factor 3 pads 3 symmetrical pixels in the boundaries of each feature map. Batch normalization (BN) is adopted right after each convolution and before the activation function. We initialized the weights as in [15], and Adam is used as the minimizing function with a mini-batch size of 32. The learning rate starts from 0.1 and is divided by 10 when the error plateaus. We use a weight decay of 0.001 and a momentum of 0.9. The dropout is not used in the training phase.

C. JOINT STRATEGY

Since the input of the neural network is an image patch, the input image at the test stage also needs to be divided into small patches. Then, the output surface normal patches are required to be jointed together. If the surface normal patches

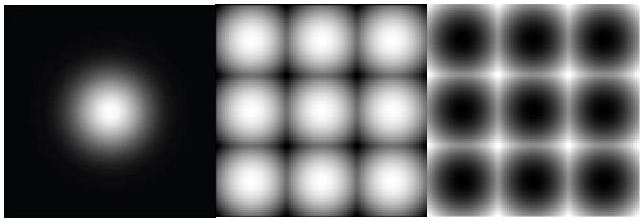

FIGURE 2. Human eye perception model and the joint templates (From left to right: human eye perception function, positive joint template and negative joint template.)

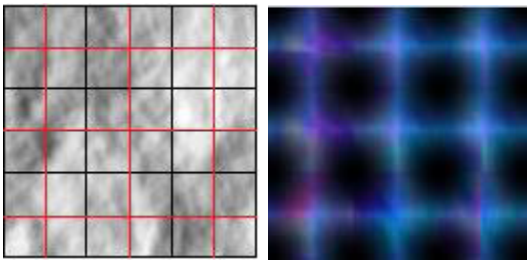

FIGURE 3. Joint strategy for the surface normal patches (The left shows the dividing strategy, and the right shows the joint strategy.)

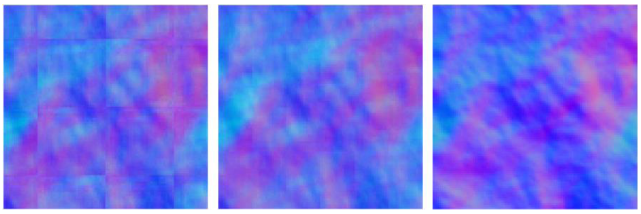

FIGURE 4. Estimated surface normal (On the left is the directly jointed surface normal, in the middle is the weighted fused surface normal, and on the right is the ground truth.)

are directly jointed, the restored surface normal may cause an annoying artifact boundary, shown in the left image in Fig. 4. There are two methods to deal with this problem: symmetrical padding and zero padding. We used the symmetrical padding strategy. Although this strategy is good for an image with a rich texture, it is not useful for a surface normal with a high continuity. To reduce the artifact boundary, we propose a new joint strategy based on the perception model of human eyes [23]. Physiological research has shown that the human visual system has characteristics of rapid perception, and the Gabor function can well simulate these mechanisms of human eye perception. The 2D Gabor function is defined as follows:

$$Gabor(u, v, \delta) = \frac{1}{2\pi\delta^2} e^{-\frac{u^2+v^2}{\delta^2}} \tag{2}$$

where δ is the variance of the Gabor function, and (u, v) represents the direction of the Gabor function. The left image in Fig. 2 shows the 2D Gabor function. To reduce the artifact boundary of patches, such as those shown in the left image in Fig. 4, we produced positive and negative joint templates along the seams of the patches, as shown in the last two images in Fig. 2. Then, the operation of dividing patches, which will cut up along the black border and the red border, respectively, as shown in the left image in Fig. 3, is performed twice for one input image. Each patch is put in SFResNet to obtain the estimated patch surface normal. According to the above dividing principle, the estimated patch surface normals will be jointed to obtain two surface normal images. The surface normal jointed along the black border is defined as n_1, and the surface normal jointed along the red border is defined as n_2. The new surface normal vector is generated by the weighted sum of n_1 and n_2 according to the following equation:

$$n_{new} = w_1 n_1 + w_2 n_2 \tag{3}$$

where w_1 represents the positive template, w_2 represents the negative template ($w_1 + w_2 = 1$), and the weighted sum is determined according to the principle of Fig. 3. The weighted fused n_{new} is shown in the middle image of Fig. 4. The left image in Fig. 4 is the surface normal that is directly jointed using patch surfaces, and the right image is the ground truth of the input image. As shown, n_{new} is more similar to the ground truth visual effects.

IV. DEEP RESIDUAL NETWORK WITH SPARSE FEEDBACKS

In our opinion, a very deep network architecture requires a huge training set; however, for many computer vision tasks, a large number of training samples are not easy to obtain. Nevertheless, small training samples can be easily constructed.

To verify the effectiveness of the proposed SFResNet, we evaluated our proposed model and method on Photex texture database. The image in the Photext database is monochrome. Many intensity images in different lighting directions were captured in this database, and rock images with resolution of 512×512 were used in these experiments. There are ten types of rock images, which include *aab*, *aaf*, *aai*, *aaj*, *aam*, *aan*, *aao*,*aap*, *aar* and *aas*. We divided each rock image into 16 pieces (each piece had a resolution of 128×128).

Since the input image size of SFResNet is small, we divided each image piece into 32×32 patches. Then, we could obtain 2685 patches; 2560 were used for training, and 125 were used for testing.

Based on the joint strategy, SFResNet can receive images of any size, which must be greater than or equal to 32×32. Figure 5 shows the estimated surface normals for an input image size of 128×128. The image in the first column is the input image, the second column shows the directly jointed surface normals, the third column shows the weighted fused surface normal, and the last column shows the ground truth. From the reconstructed results, we observed that the weighted fused surface normal reduced the artificial Mosaic trace at the seam of patches, and more visually pleasing output results were produced compared with that of the surface normal that was directly jointed.

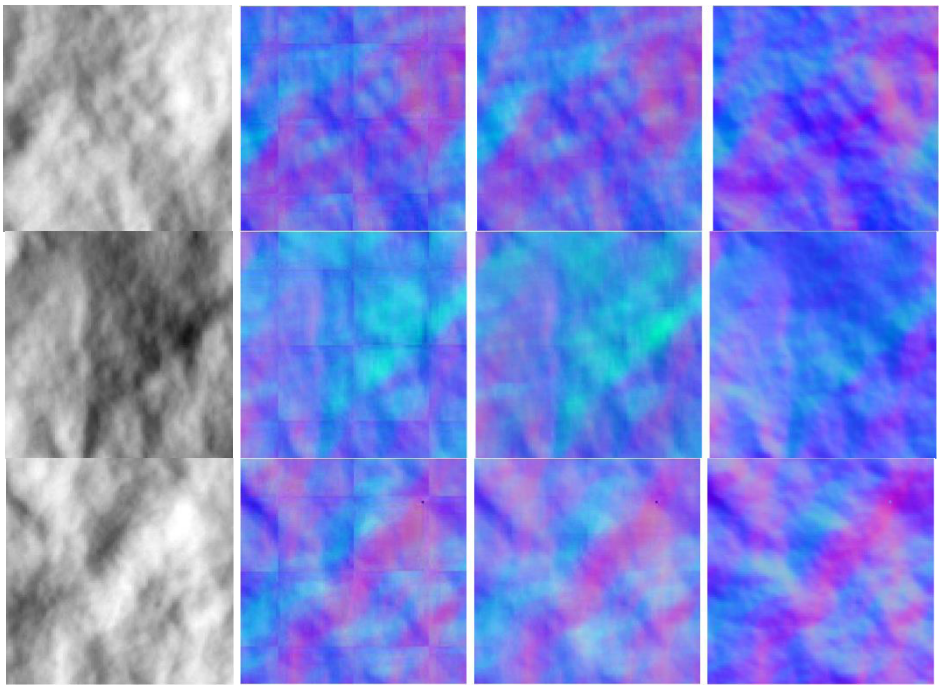

FIGURE 5. Reconstructed surface normal (The first column shows the input images, the second column shows the directly jointed surface normals, the third column shows the weighted fused surface normal, and the last column shows the ground truth.)

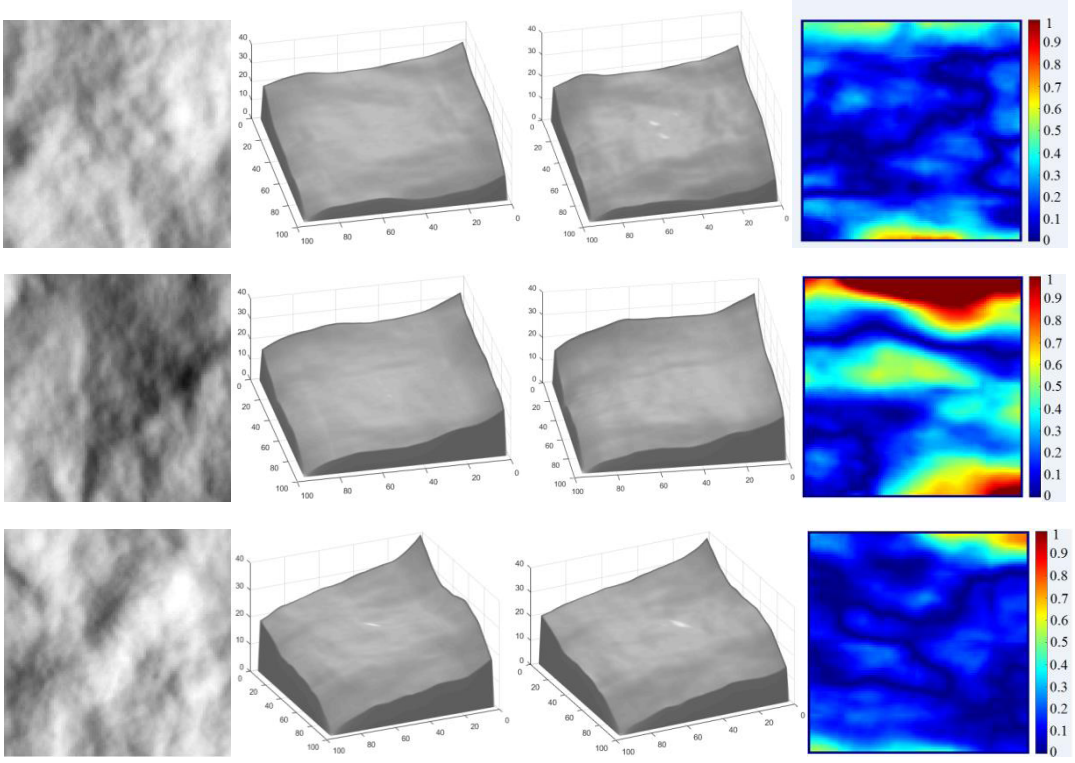

FIGURE 6. Reconstructed height maps (The first column shows the input images, the second column shows the reconstructed 3D shapes integrated with the weighted fused surface normals, the third column shows the 3D shapes of the ground truth, and the last column shows the relative error maps.)

Figure 6 shows the reconstructed height maps, which were integrated [24] from the reconstructed surface normals in Fig. 5. The first column shows the input images of SFResNet; the second column shows the reconstructed 3D shapes, which are integrated from the weighted fused surface normals, and we call them fused 3D shapes; the third

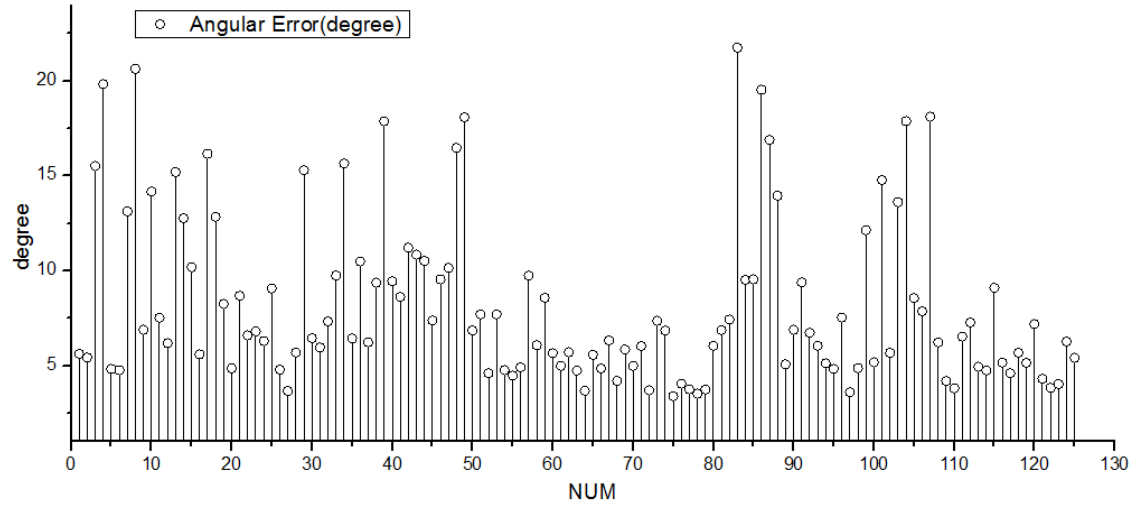

FIGURE 7. Angular error statistics for the tested images.

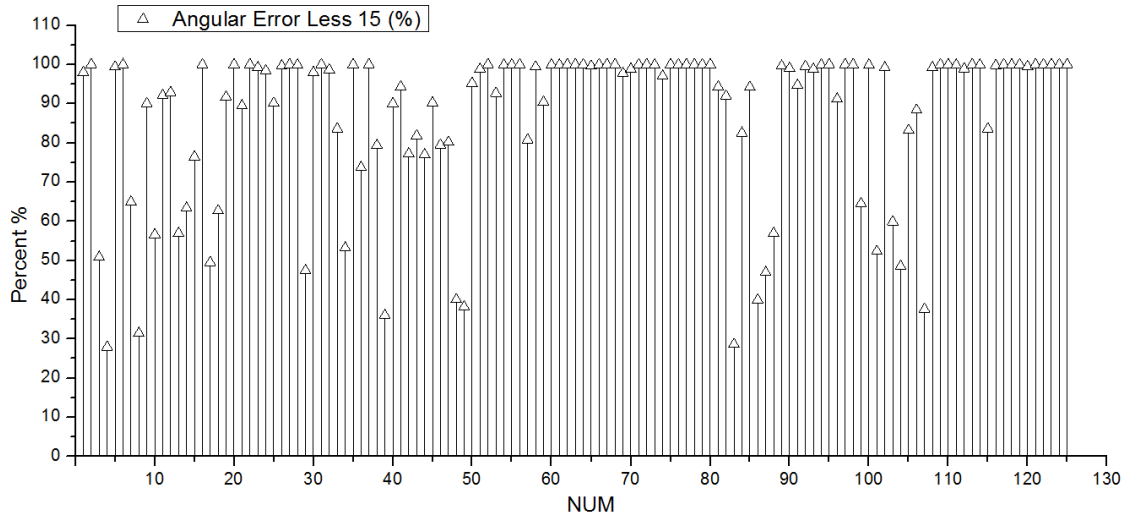

FIGURE 8. Angular error statistics of less than 15 degrees for the tested images.

column shows the 3D shapes for the ground truth; the last column show the relative error maps between the fused 3D shapes and the ground truth. The relative error map is defined as:

$$err\_map = \frac{|z(x,y) - z_{gt}(x,y)|}{z_{gt}(x,y)}|_{x,y\in\Omega} \times 100\% \qquad (4)$$

where x and y represent the Cartesian coordinates of one pixel in the input image; Ω represents the total pixels in each image. $Z(x,y)$ and $Z_{gt}(x,y)$ represent the reconstructed 3D shape and the ground truth, respectively. Because the error value is small, the relative error map computed according to equation (4) looks slightly dark. To more clearly show the error map, the intensity of the error map in Fig. 6 was magnified ten times. The number on the right of the error map represents the percentage distribution of the error map; of course, these numbers were also magnified ten times.

The height error map in Fig. 6 consists of two parts: one is the normal vector estimation error, and the other is the height error caused by the integration of the surface normal. To measure the estimated accuracy of SFResNet, we only show the error statistics for the estimated surface normal of SFResNet, and the angle error was used as the criterion judgment, which is defined as:

$$angular\_err(n, n_{gt}) = \arccos \frac{n \cdot n_{gt}}{\|n\| \left\|n_{gt}\right\|} \qquad (5)$$

where arccos represents the anticotangent function of the estimated surface normal and that of the ground truth. If the difference between n and n_{gt} is small, there will be a small angular error. Figures 7 and 8 show the statistics for the angular errors for the 125 tested images. The vertical coordinate of Fig. 7 represents the degree of angular error of the estimated surface normal, and the horizontal coordinate represents the number of tested images. As shown in Fig. 7, most of the

angular error is less than 15 degrees. To show a percent of less than 15 degrees, we re-ran the statistics, as shown in Fig. 8. The meaning of the horizontal coordinate is same as with Fig. 7, and the vertical coordinate of Fig. 8 represents a percent of less than 15 degrees for the *angular_err*. As shown by the distribution in Fig. 8, an angular error of less than 15 degrees is high and robust except for in a few images. The reason for the large error for these few images may be that our training data are not sufficient; however, the correctness of most of the data verifies the rationality of our network model.

V. CONCLUSION

In this study, we designed and trained a deep residual network with sparse feedbacks for 3D reconstruction from one image. The surface normal of the input image was first estimated by using the proposed SFResNet, and then the 3D surface shape of the object was integrated from the estimated surface normal. By using the method of estimating the surface normal, we produced a smoother surface height map than that produced via direct estimation of the height map.

Our network can receive a random size image by using a segmentation strategy for the input image, and the proposed joint strategy can seamlessly connect the segmented small patches. In addition, the forward sparse feedback improves the convergence speed and the training stability of the network model. Hence, visually pleasing output results can be produced for different input images.

REFERENCES

[1] G. Wang and J. Cheng, "Three-dimensional reconstruction of hybrid surfaces using perspective shape from shading," *Optik*, vol. 127, no. 19, pp. 7740–7751, 2016.

[2] X. Wang, L. Gao, J. Song, and H. Shen, "Beyond frame-level CNN: Saliency-aware 3-D CNN with LSTM for video action recognition," *IEEE Signal Process. Lett.*, vol. 24, no. 4, pp. 510–514, Apr. 2017.

[3] L. Bottou *et al.*, "Comparison of classifier methods: A case study in handwritten digit recognition," in *Proc. Int. Conf. Pattern Recognit.*, vol. 2, Oct. 1994, pp. 77–82.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates, 2012, pp. 1097–1105.

[5] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1–9.

[6] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, 2017.

[7] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[8] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[9] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[10] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification?" *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5017–5032, Dec. 2015.

[11] Y. Lecun *et al.*, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Dec. 2014.

[12] Y. Sun, M. Jian, X. Zhang, J. Dong, L. Shen, and B. Chen, "Reconstruction of normal and albedo of convex Lambertian objects by solving ambiguity matrices using SVD and optimization method," *Neurocomputing*, vol. 207, pp. 95–104, Sep. 2016.

[13] Y. Sun, X. Zhang, M. Jian, S. Wang, Z. Wu, Q. Su, and B. Chen, "An improved genetic algorithm for three-dimensional reconstruction from a single uniform texture image," *Soft Comput.*, vol. 22, no. 2, pp. 477–486, 2018.

[14] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1440–1448.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. ICCV*, 2015, pp. 1026–1034.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[17] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.

[18] V. Jain and H. S. Seung, "Natural image denoising with convolutional networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates, 2008, pp. 769–776.

[19] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates, 2012, pp. 341–349.

[20] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[21] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[22] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2392–2399.

[23] Y. Sun, Q. Su, X. Zhang, and P. Liu, "The main lighting direction estimation for the uniform texture images," in *Proc. IEEE Asia–Pacific Signal Inf. Process. Assoc. Summit Conf.*, Dec. 2015, pp. 769–772.

[24] B. K. Horn, "Height and gradient from shading," *Int. J. Comput. Vis.*, vol. 5, no. 1, pp. 37–75, 1990.

[25] X. Zhou, G. Zhong, L. Qi, J. Dong, T. D. Pham, and J. Mao, "Surface height map estimation from a single image using convolutional neural networks," *Proc. SPIE*, vol. 10225, p. 1022524, Feb. 2017.

[26] G. Trigeorgis, P. Snape, I. Kokkinos, and S. Zafeiriou, "Face normals 'in-the-wild' using fully convolutional networks," in *Proc. IEEE Computer Vis. Pattern Recognit.*, Jul. 2017, pp. 340–349.

[27] E. W. Y. So, S. Michieletto, and E. Menegatti, "Calibration of a dual-laser triangulation system for assembly line completeness inspection," in *Proc. IEEE Int. Symp. Robotic Sensors Environ.*, Nov. 2012, pp. 138–143.

[28] F. Liu, C. Shen, and G. Lin, "Deep convolutional neural fields for depth estimation from a single image," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5162–5170.

[29] D. Eigen and R. Fergus, "Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture," in *Proc. ICCV*, Dec. 2015, pp. 2650–2658.

YUJUAN SUN received the Ph.D. degree from the Department of Computer Science and Technology, Ocean University of China. She is currently an Associate Professor with the College of Information and Electrical Engineering, Ludong University, China. She has published more than 20 journal and international conference papers (indexed by SCI, EI, and ISTP). Her research interests include 3-D shape reconstruction from images, pattern recognition, and machine learning.

MUWEI JIAN received the Ph.D. degree from the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, in 2014. He joined the Department of Computer Science and Technology, Ocean University of China, as a Lecturer, in 2015. His current research interests include image processing, wavelet analysis, 3-D multimedia analysis, and human face hallucination/recognition. He has served as a reviewer for several international SCI indexed journals, including *Pattern Recognition*, *Computers in Industry*, *The Imaging Science Journal*, *Machine Vision and Applications*, and *Multimedia Tools and Applications*. He holds three granted national patents and has published over 30 papers in refereed international leading journals/conferences, such as the IEEE TRANSACTIONS ON CYBERNETICS, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, *Information Sciences*, *Pattern Recognition*, *Signal Processing*, ISCAS, ICME, and ICIP.

XIAOFENG ZHANG received the master's degree from the Lanzhou University of Technology in 2005 and the Ph.D. degree from Shandong University in 2014. He is currently an Associate Professor at the College of Information and Electrical Engineering, Ludong University, China. He has published approximately 30 papers in international journals and conferences. His research interests include medical image segmentation, pattern recognition, and among others.