Multiple active stereo systems calibration method based on Neural SDF using DSSS for wide area 3D reconstruction

Kota Nishihara[1*], Ryo Furukawa[2], Ryusuke Sagawa[3], and Hiroshi Kawasaki[1]

[1] Kyushu University, Fukuoka, Japan
[2] Kindai University, Hiroshima, Japan
[3] AIST, Tsukuba, Ibaraki, Japan

Abstract. One of the practical methods for general 3D shape reconstruction is active stereo, which involves capturing images of scenes illuminated by structured light and reconstructing the scenes from the images. Active stereo has the advantage of being able to achieve high-density reconstructions even for texture-less scenes. However, it requires finding dense correspondences between the captured image and the projected pattern, making it difficult to capture the scene using multiple systems simultaneously where patterns are overlapped each other. A simple solution is to switch the patterns on-and-off synchronously; however, this naturally decreases exposure time in inverse proportion to the number of systems, resulting in a lower signal-to-noise ratio (SNR). In addition, retrieving relative poses between a projector and a camera of active stereo systems as well as relative poses between multiple systems is another challenge. To address these challenges, we propose a technique based on neural signed distance field (Neural-SDF) using Direct Sequence Spread Spectrum (DSSS). DSSS is the latest technique widely used in the field of communications for multiplexing multiple signals and separating them. We propose a novel method to utilize DSSS to project multiple structured light patterns onto the object, where the overlapped patterns are efficiently separated. To calibrate the relative poses between projectors and cameras in multiple sets of active stereo systems, a differential renderer based method using Neural-SDF is proposed. Such an approach has not yet been explored yet for active stereo systems. In the experiments, it was proved that our technique worked successfully by both qualitative and quantitative evaluations using real sensors and objects.

Keywords: Structured light · camera and projector calibration · DSSS · Neural SDF.

1 Introduction

In recent years, 3D shape reconstruction has been used in various fields. One of the common methods for 3D shape reconstruction is the multi-view stereo

* Corresponding author. Email address: nishihara.kouta.869@s.kyushu-u.ac.jp

(MVS) using only RGB cameras. In MVS, the object is captured by two or more cameras and the 3D shapes are reconstructed by using the correspondences between images, which makes it difficult to restore a high-density shape of an object when there is a small texture. On the other hand, there is an active stereo method in which one of the cameras is replaced by a projector. In the method, a high frequency pattern is projected onto the object, and shape restoration is performed by finding the corresponding points using the projected pattern. Since image features are added to the object by the projected pattern, even if the object has only a small texture, high-density restoration can be achieved.

One issue of active stereo systems is that if multiple projectors are used simultaneously, patterns are overlapped each other and it becomes difficult to separate them. A simple solution is to switch the patterns synchronously; however, this naturally decreases exposure time in inverse proportion to the number of projectors, resulting in a lower signal-to-noise ratio (SNR). In this paper, we propose a technique based on Direct Sequence Spread Spectrum (DSSS), commonly used in radio communication to separate multiplexed signals. By using the technique, it not only enables the separation of each structured light, but also eliminates the influence of external light, which significantly decreases a SNR.

Another issue of active stereo systems is that calibration of extrinsic parameters between the projector and the camera. In addition, rigid transformation between the systems are also challenging. To solve the issues, novel calibration technique to simultaneously estimate the extrinsic parameters of multiple projector-camera systems as well as rigid transformation between them are proposed. In our technique, neural singed distance field (Neural-SDF) is used, where differences between captured images and rendered images are minimized.

In the experiments, it was proved that our technique worked successfully by both qualitative and quantitative evaluations using real devices and objects. In summary, we provide three main contributions of this approach.

1. We propose a technique based on DSSS to separate multiple patterns overlapped each other from multiple viewpoints as well as improve the SNR.
2. We propose a novel approach to calibrate multiple projectors and cameras of active stereo systems with high accuracy using Neural-SDF without use of a special calibration instruments.
3. Comprehensive experiments are conducted by using actual devices and objects having successfully demonstrated the effectiveness of the proposed method.

2 Related Works

2.1 Dense reconstruction by active stereo techniques

Active stereo is a 3D shape measurement technique, where a pattern is projected from a projector onto the object, and the image taken by a calibrated camera is used to recover the 3D shapes by stereo algorithm [2,10]. Among them, one-shot scan, which uses a spatial-coded pattern to measure the shape by a single image, draw wide attention, since it is suitable for measurement of moving objects [12, 19, 20, 29, 31] or by moving cameras [21]. A key technique of one-shot scan is efficient correspondence finding between the projection pattern and the projected pattern on the object surface in the captured image [18,26,29]. Recently, various methods have been developed for detecting the correspondences in stereo pairs using deep learning [4,27,34,35], however, they are basically for passive stereo and achieving a dense correspondence retrieval for active stereo is also a challenging problem, since their features are almost identical and repetitive which makes it difficult to distinguish with.

2.2 Calibration of projector and camera system

To calibrate the extrinsic parameters between a projector and a camera, a typical approach is to use a calibration object, on which patterns are projected [15, 32]. These methods require calibration in advance and the relative position between the projector and the camera should be fixed each other, which significantly decreases the usability of the system, compared to passive systems where those parameters can be self-calibrated during the scan, known as SLAM or SfM [3,13,14,17,28,36]. Recently, Furukawa *et al*. proposed a SLAM method for active stereo systems using GCN-based estimation of 2D correspondence points performing auto-calibration [6]. However, it requires several unique feature points to calibrate the system and, therefore, cannot be applied to general patterns. In contrast, our technique minimizes photometric loss using a volumetric differentiable renderer, which does not have such limitations.

2.3 Differentiable renderer

Differentiable renderer is a technique to render CG images using a scene or camera parameters where gradient-based optimization w.r.t. the parameters can be processed. Mesh-based differentiable renderers [9, 11, 16, 25] have been proposed to optimize various scene parameters such as geometry, illumination, textures, or materials and intensively researched, however, point clouds or occlusions cannot be handled. Recently, voxel-based differential renderers draw much attention because of success of handling more general conditions, such as NeRF [22]. Especially SDF based neural representation is promising to reconstruct shape [30]. In this paper, we use a SDF based volumetric differentiable renderer.

3 Overview

3.1 System configuration

The system configuration is shown in Fig.1(a). The measurement is performed using multiple sets of projector-camera systems. During scan, the projectors and cameras are all fixed each other. A laser pattern projector is used to projects structured light pattern, which is grid-like shape formed by diffractive optical element (DOE), onto the measurement target, and then, scenes are captured by

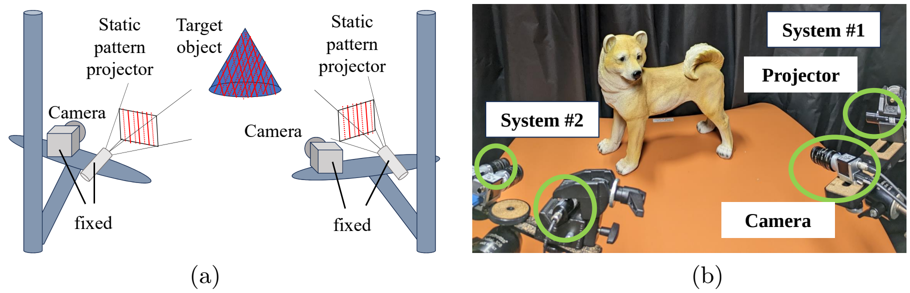


Fig. 1: Our measurement system. (a) Schematic configuration of the system. (b) Real capturing system where two projector-camera systems are installed.

the cameras. In the system, infrared is used for pattern projector not to interfere RGB colors. The projectors are capable of blinking at a high frame-rate, 90fps in our system, so that DSSS can be applied to moving objects. All the systems are synchronized to apply DSSS. Real system we built for experiments are shown in Fig.1(b), where two sets of a projector and a camera systems were installed.

3.2 Active stereo system using single pattern projector

In this paper, we use one-shot active stereo system, which reconstructs 3D shapes from a single captured image using spatial encoding. We adopt the state-of-the-art algorithm which utilizes deep learning to obtain dense correspondences between the captured image and the projected pattern image at the pixel level [5].

The reconstruction process is shown in Fig.2. In the method,[4] initially, the phase information of the pixels in the captured image is detected using U-net(Fig.2(a)). Segmentation is performed based on the estimated phase, creating a grid structure(Fig.2(b)). Additionally, the nodes of the projected pattern carry five types of features, and these features are estimated from the captured image using U-net(Fig.2(c)). The detected grid structure and code information of nodes represent a graph(Fig.2(d)), which are fed into GCN (Graph Convolutional Network)(Fig.2(e)) to obtain node-wise correspondence mapping(Fig.2(f)). Then, dense correspondence is achieved through interpolation between nodes using phase information, which is output as x, y-pixel-wise projector-coordinate images(Fig.2(g)). Finally, 3D shapes are reconstructed from dense correspondences using triangulation algorithm(Fig.2(h)).

3.3 Overview of the entire scanning process

The entire flow of wide area 3D reconstruction and calibration process are shown in Fig.3. Note that our technique can calibrate the system using arbitrary-shaped objects for each frame; however, in real experiments, it is preferable to use

[4] The followings are a brief explanation of the method [5], however, it is not necessary to understand our method, so readers who are familiar with active stereo can skip the rest of this section.

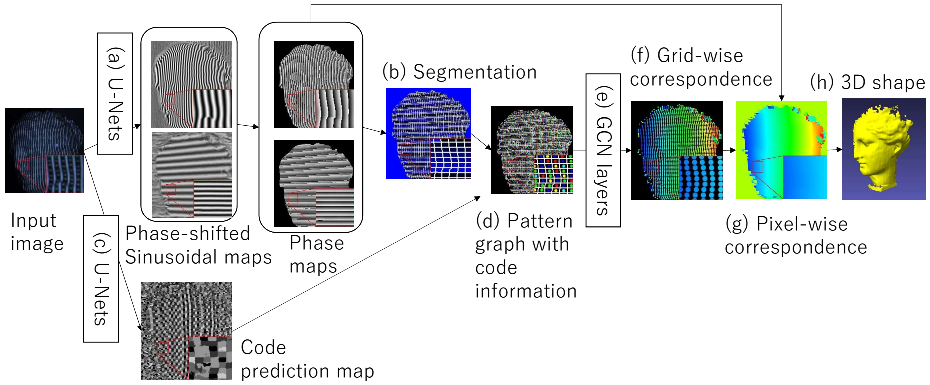


Fig. 2: Flow of one-shot dense active-stereo system used in out method.

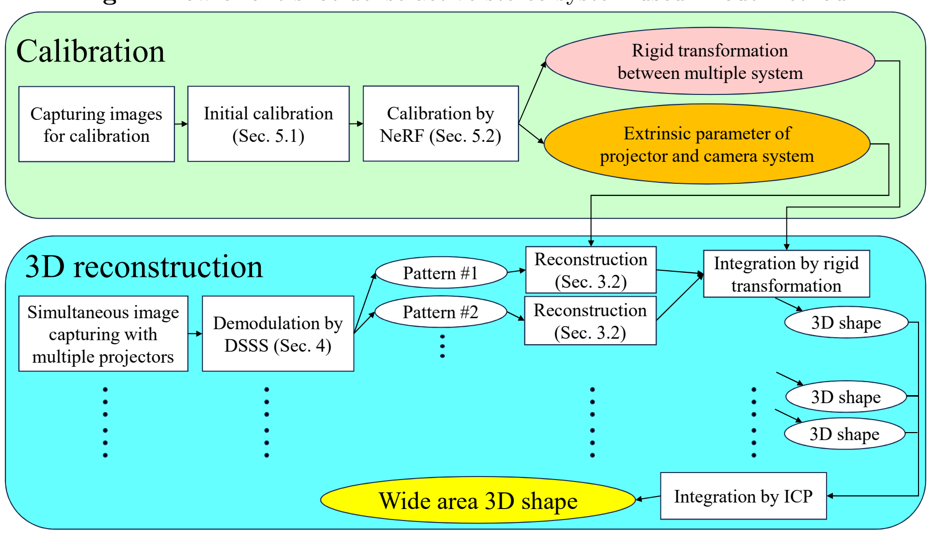


Fig. 3: Overview of the 3D object scanning process of wide area

simple-shaped objects with a certain number of inputs to achieve better results. Therefore, we typically adopt a two-step approach consisting of calibration and reconstruction steps.

In first step, the calibration of the system is conducted. Auto-calibration for one-shot scan [6] is performed to obtain initial values of extrinsic parameters. Then, simultaneous calibration of both extrinsic and rigid transformation based on NeRF is performed using the initial parameters and the corresponding points obtained by U-net and GCN, which are retrieved in the process of Sec.3.2.

In the next step, the target object on which the multiple patterns are projected is captured from multiple viewpoints to reconstruct a wide area. Although multiple patterns are projected onto the object, only a single pattern should be extracted for 3D reconstruction. The separation is achieved by DSSS. Using

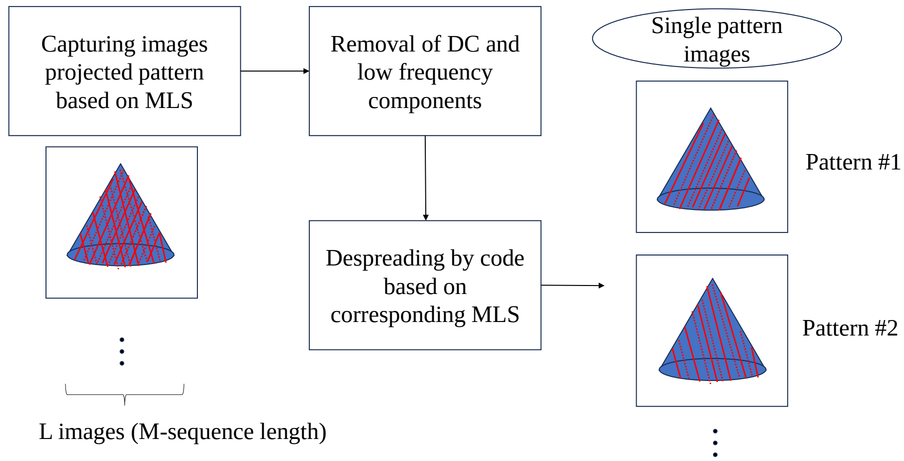


Fig. 4: Demodulation of multiple patterns for one-shot scan using DSSS.

these separated images as the input, a dense 3d reconstruction is performed by oneshot scan algorithm explained in Sec.3.2. The final wide areal 3D shape is obtained by integrating the reconstructed shapes from multiple viewpoints by aligning them using the estimated rigid transform parameters.

4 Separation of overlapped patterns using DSSS

DSSS is a signal modulation technique used in radio communications that can reduce the effects of noise and interference signals. In DSSS, the signal is spread by multiplying the transmitted signal by a spreading code. We apply DSSS to modulate light blinking timing on-and-off not only for separating overlapped patterns, but also for improving SNR. Fig.4 shows the flow of demodulation by DSSS in our method. We used a spreading code based on maximal-length sequence (MLS) [7] which is a pseudo-noise sequence and its length is $2^N - 1$ for DSSS. A circular-shifted sequence of MLS is also a type of MLS. We convert the value of -1 in MLS to 0 and use 1 for laser to project pattern and 0 for no pattern. For a signal s projected by the projector, DSSS multiply it by a spreading code S as follows:

$$M(t) = S \circ s(t), \tag{1}$$

where $M(t)$ is the signal we projected and the operator $\circ$ means an entry-wise product. Then, the projected signal is received with interfered by noise:

$$M'(t) = M(t) + n(t), \tag{2}$$

where $M'(t)$ is the received signal and $n(t)$ is the noise. Then, the signal need to be despread by multiplying despreading code S':

$$s'(t) = M'(t) \circ S'. \tag{3}$$

During the despreading, we get the original signal and spread the noise to remove it's interference. In our system, since we used two sets of the active stereo system, we set $N = 3$, and thus, we get 0,0,1,0,1,1,1 for projector #1 and 0,1,0,0,1,1,1,0 for projector #2 with one bit shift. Multiple cameras simultaneously capture the same number of images for the length of MLS in accordance with the blinking timing of the projectors. The captured images contain multiple projected patterns overlapped and the external light sources (noises), which sometimes significantly decrease the SNR. As the correlation between different MLSs is small, we can demodulated single pattern from the captured images by multiplying spreading code based on used MLS, and DC and low frequency components can be removed by applying a high-pass filter. After the demodulation, as only a single pattern is observed in the image, the one-shot scan reconstruction technique can be applied without any modification. As a result, shapes are reconstructed for each device.

5 Auto calibration by Neural SDF

5.1 Initial auto-calibration for static pattern for oneshot scan

Optimizing parameters using Neural SDF requires reasonably accurate initial values, therefore, initial parameters are obtained through recently proposed auto-calibration technique [6]. With the technique, several images are captured where static patterns are projected onto arbitrary object and dense correspondences are obtained by code information estimated by U-net and GCN. Then, RANSAC algorithm is applied to remove outliers from estimated correspondence points by sampling eight correspondence points to obtain the extrinsic parameter between a camera and a projector.

5.2 Optimization by volumetric differential renderer using photometric loss

To improve calibration accuracy, neural shape representation using SDF as same as NeuS [30] and a volume renderer-based optimization algorithm is proposed as shown in Fig.5. We use NeuS instead of NeRF in our method because surfaces of objects reconstructed by NeuS are more accurate than those of NeRF. Moreover, rendered results of a ray in NeRF can be a weighted average of multiple high-density regions that are separated, which does not occur in active stereo systems. For NeuS, rendered results of a ray is basically a weighted average of single region around a zero-crossing of SDF. Although arbitrary-objects can be used in the calibration step, in order to stabilize the optimization, we use a three-plane corner shape as the measurement target. Then, in the reconstruction step, the calibrated parameters are used as shown in Fig.3. We capture it from multiple viewpoints, and x, y-projector-coordinate images are obtained by estimating U-net and GCN. To utilize information from active pattern projection, we render projector-coordinate images by differential volume renderer using the

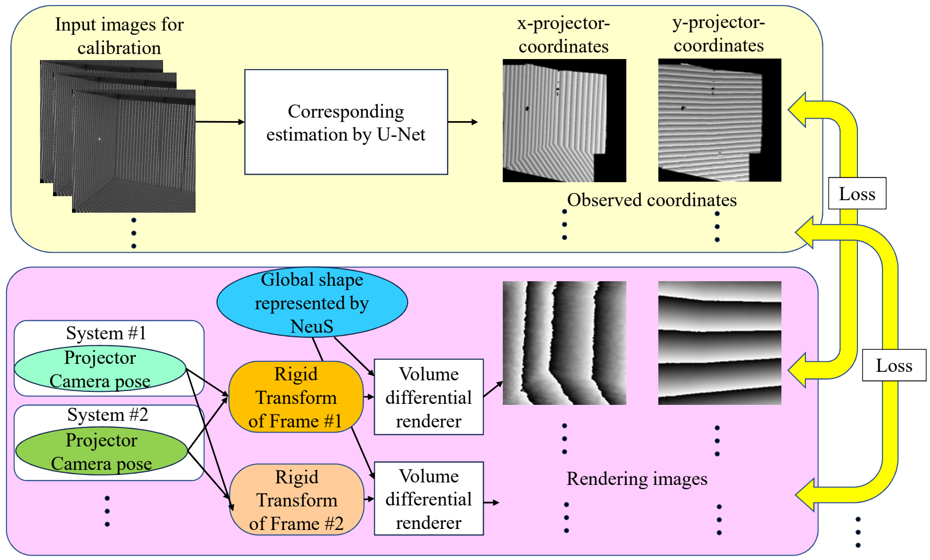


Fig. 5: Flow of optimizing camera and projector poses based on NeRF

neural scene representation by MLP where cameras and projectors poses are set as the learnable parameters. Detailed process of each step is as follows.

First, x-projector-coordinate image and a y-projector-coordinate image are obtained from captured image for calibration, where a combination of a x-projector-coordinate image and a y-projector-coordinate image represent a 2D-to-2D mapping:

$$\boldsymbol{H_f} \ : \mathbb{R}^{\mathbf{2}} \mapsto \mathbb{R}^{\mathbf{2}} \, ; \boldsymbol{r} \mapsto \boldsymbol{q} \tag{4}$$

from camera pixels r to projector pixels q. H_f represents all the observed information obtained at the f-th frame.

We use c, which maps a 3D point $\mathbf{p}$ to '2D projector coordinates' in the similar way as 'projection mapping' used in CG. Thus, $\mathbf{c} : \mathbb{R}^{\mathbf{3}} \rightarrow \mathbb{R}^{\mathbf{2}}$ is defined as

$$\mathbf{c}(\mathbf{p}) = \frac{1}{-z'} \begin{bmatrix} \alpha_x x' \\ \alpha_y y' \end{bmatrix} + \begin{bmatrix} \beta_x \\ \beta_y \end{bmatrix}, \text{ for } \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \mathbf{R}_{wp} \ \mathbf{p} + \mathbf{t_{wp}}, \tag{5}$$

where $\mathbf{R}_{wp}$ and $\mathbf{t}_{wp}$ are rotation and translation of the world-to-projector coordinate transformation.

The surface $\mathcal{S}$ is represented as a zero level set of SDF in the same way as NeuS [30].

$$\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^{\mathbf{3}} | \mathbf{f}(\mathbf{x}) = \mathbf{0}\}. \tag{6}$$

We render a pixel correspondence maps by using the following rendering method. A ray from a camera can be represented by camera optical center $\mathbf{o}$ and ray direction $\mathbf{v}$, as $\{\mathbf{p}(t) = o + t\mathbf{v} | t \geq 0\}$. The rendered pixel value is as follows:

$$C(\mathbf{o}, \mathbf{v}) = \int_0^{+\infty} \mathbf{w}(t)\mathbf{c}(\mathbf{p}(t))dt. \tag{7}$$

Note that $\mathbf{c}(\mathbf{p})$ is not a radiance-field color vector. Instead, we use xy-coordinates of a projector represented as Eq.(5). Also, by using the weighting function $\mathbf{w}(t)$ similar to NeuS, we can selectively put high weights around zero-level surfaces of the SDF. So, by using $\mathbf{c}(\mathbf{p})$ instead of a radiance-field color vector, we can render "correspondence maps," which is a similar calculation with projection mapping in CG.

For the Neural SDF, we use hash-grid-encoding representation [23]. Since we do not use radiance-field color vectors, the output of the field function is only the SDF function. We optimize the Neural SDF and all the camera/projector poses by minimizing discrepancies between the projector-coordinate images and the target coordinates obtained from the observed patterns.

For loss functions for the optimization, we utilize the sum of L1 losses for the projector-coordinate images for all the frames. To avoid effects of outliers affecting the optimization, we mask the image regions where the projector-camera correspondences are not obtained. Also, we add Eikonal loss [8] along the sampled rays to regularize the SDF.

6 Experiment

6.1 Evaluation of demodulation using four projector camera systems

We conducted an experiment to demonstrate the effectiveness of pattern separation for one-shot scan using DSSS. We projected patterns from four directions using four sets of camera-projector systems. Each of the four laser pattern projectors captured seven images based on MLS: [0,0,1,0,1,1,1], [0,1,0,1,1,1,0], [1,0,1,1,1,0,0] and [0,1,1,1,0,0,1]. The captured images and the demodulation results are shown in Fig.6. It is confirmed that overlapped patterns are successfully separated by our DSSS based algorithm. These demodulated images allow for shape reconstruction from multiple viewpoints.

In DSSS, since the pattern is redundant to increase SNR, it is important to confirm the effectiveness of our method. A comparison was conducted under a strong external illumination with and without DSSS. In the experiment, target objects are illuminated by a video projector as a strong external light source. Fig.7 shows separation results with and without DSSS. To make comparison fair, total exposure times for both methods are adjusted, such as 18ms * 2 frames = 36ms for single pattern alternating projection, whereas 5ms * 7 frames = 35ms for DSSS projection. From the separation results, we can confirm that DSSS result is better than naive method. PSNR were 25.60 for naive method and 31.36 for DSSS, respectively.

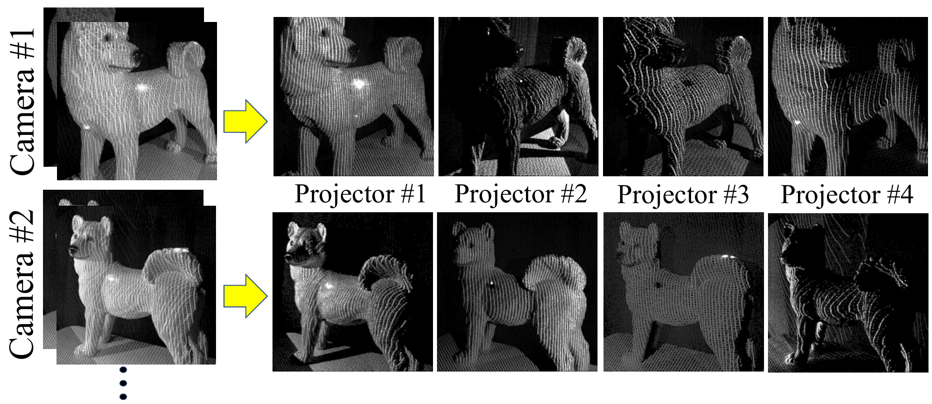


Fig. 6: Scene scanning using four devices and separation results.

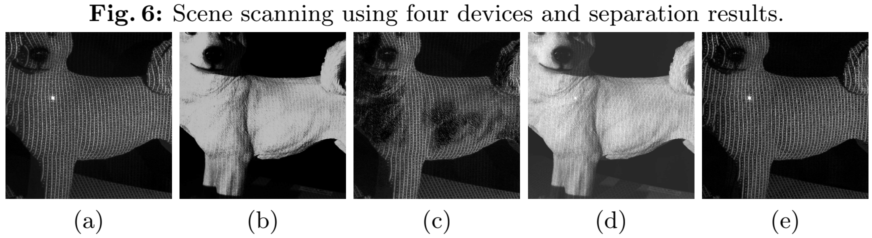


Fig. 7: Overlapped pattern separation results. (a) ground truth, (b) single pattern projection with 18ms exposure time, (c) normalization of (b) after subtracting background image without pattern projection, (d) two patterns projection with 5ms exposure time, and (e) separation results using DSSS from 7 images of (d). We can clearly confirm that our result (e) is similar to ground truth (a) than result of simple method (c).

Next, we conducted phase and code estimation using U-net as shown in Fig.2(a) and (c). The results are depicted in Fig.8. When DSSS is not applied, certain portions of the phase and code estimation failed due to the influence of the external light source. However, by utilizing DSSS, the impact of the external light source was effectively decreased and it is confirmed that severe noises are successfully eliminated as shown in close up view in Fig.8(bottom). From the results, it can be confirmed that code estimation algorithm gained more benefit from DSSS, since the image features were more easily affected by noise.

In addition, we confirmed the improvement of the projector-coordinate estimation, which was an output of GCN(Fig.2(e)) using both phase and code images as the input and reconstructed shapes. The results are shown in Fig.9, where the areas illuminated by external light sources were severely deteriorated if DSSS was not applied. The ratio that has error of 1 pixel or more from the reference in the projector coordinates in the area was 5.5% for the image without DSSS, whereas 1.0% for the image with DSSS, and decrease of outliers in reconstructed results quantitatively confirmed the effectiveness of our method.

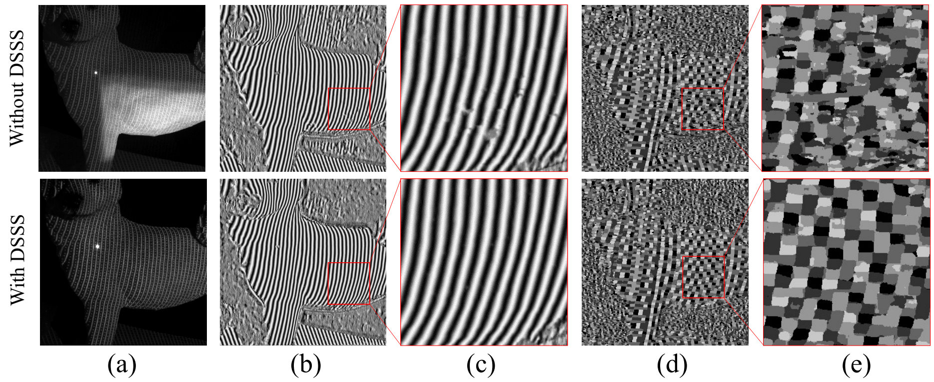


Fig. 8: Effect of DSSS. (a) Input images. Bottom image was after applying DSSS. (b) Phase estimation results by U-Net. (c) Close up views of (b). (d) Code estimation results by U-Net. (e) Close up views of (d).

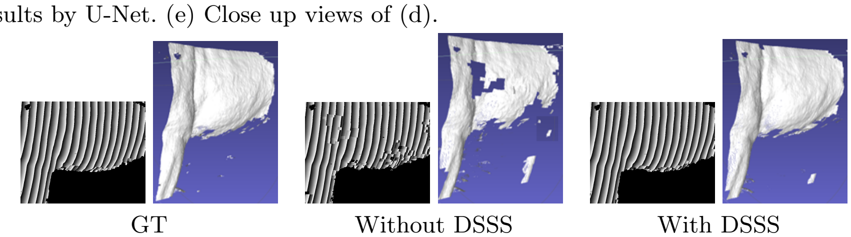


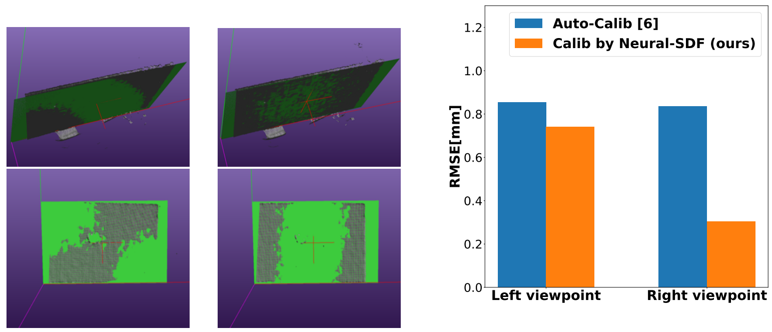
Fig. 9: Projector-coordinate images of Fig.8, cropped where strong external light was illuminated and reconstructed shapes of them.

6.2 Evaluation of calibration compared to previous methods

To evaluate the accuracy of camera and projector pose optimization using Neural SDF, we captured images of a planar board and calculated RMSE with respect to a ground truth. Fig.10(a)(b) show the reconstruction results using parameters obtained by previous auto-calibration method [6] and parameters obtained by our Neural-SDF based calibration method. Fig.10(c) shows the RMSE values for both results. Notably, the reconstruction results using Neural-SDF exhibit smaller RMSE values for both viewpoints, *i.e.*, systems #1 and #2.

The error between the two systems is measured by scanning the angles of the corners of several boxes by both viewpoints and compared the differences between them. Fig.11(a-d) show the restoration results. Compared to the results using auto-calibration method [6] with results using the parameters obtained by our method, the errors between two shapes are apparently decreased. Fig.11(e) shows the quantitative results of the improvement, confirming the effectiveness of our method.

Next, we compared our method with several previous methods by reconstructing shapes of several statues. We conducted KinectFusion [24] to obtain ground truth as shown in Fig.12(leftmost). We applied three previous methods, such as Gray-code based self-calibration method [33], hard calibration using



(a) (b) (c)

Fig. 10: Plane reconstruction results. Gray color point cloud is reconstructed shape and light green color point cloud is ground truth (GT). (a) Results using parameters of auto-calibration [6]. (b) Results using parameters of calibration based on Neural SDF (ours). (c) RMSE of (a) and (b) compared to GT.

known-shape [32] and auto-calibration based on markers imposed into the pattern [6]. Results are summarized in Tab.1 and depicted examples of reconstructed shapes are shown in Fig.12. From the results, we can confirm that our method best performed among state of the art methods.

6.3 Demonstration

Finally, we measured several objects by our multiple one-shot scanning system. Reference images of target objects are shown in Fig.13 1st column. Recovered results of each system is shown in Fig.13 2nd and 3rd column. Integrated shapes of the reconstruction results are shown in Fig.13 4th and 5th column, where effectiveness of wider scene reconstruction by using the rigid transformation parameters estimated by our Neural-SDF based calibration method are confirmed.

7 Conclusion

In this paper, we propose a wide area 3D reconstruction method by proposing pattern separation method based on DSSS as well as efficient calibration method based on Neural-SDF for multiple active stereo systems. Since DSSS is widely utilized for telecommunication purposes for enhancing SNR under severe noise condition, our proposed method is expected to separate multiple projection patterns in active stereo systems without decreasing SNR. To achieve efficient calibration of multiple oneshot scanning systems, Neural-SDF with volumetric differential renderer is applied to reduce the error among multiple systems when integrating the reconstructed results. Comprehensive experiments are conducted to prove the effectiveness of our method. In the future, we plan to increase the

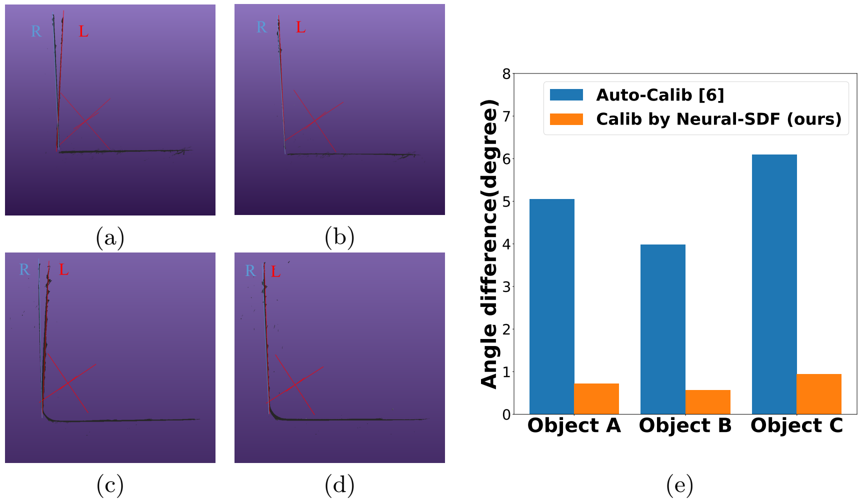


Fig. 11: Reconstruction results of several box-shaped objects. (a) Corner A using parameters estimated by auto-calibration method [6]. (b) Corner A using parameters estimated by our method. (c) Corner B of auto-calibration method [6]. (d) Corner B of our method. (e) Angle difference of the same corner scanned by system #1 and #2. The smaller the better.

GT [24] Gray-code [33] Hard-calib [32] Auto-calib [6] Ours

Fig. 12: Shape differences from ground truth of previous methods. The more dappled the pattern is, the closer the two shapes are and the smaller the error. It can be confirmed that our method achieved most spotted pattern.

Table 1: Reconstruction accuracy comparison with previous methods.

RMSE[mm]	Gray-code [33]	Hard-calib [32]	Auto-calib [6]	Neural SDF calib (ours)
Plaster dog	2.03	1.54	1.60	1.40
Ceramic hen	1.13	1.31	1.21	1.07
Plaster object	1.22	1.26	1.29	1.20

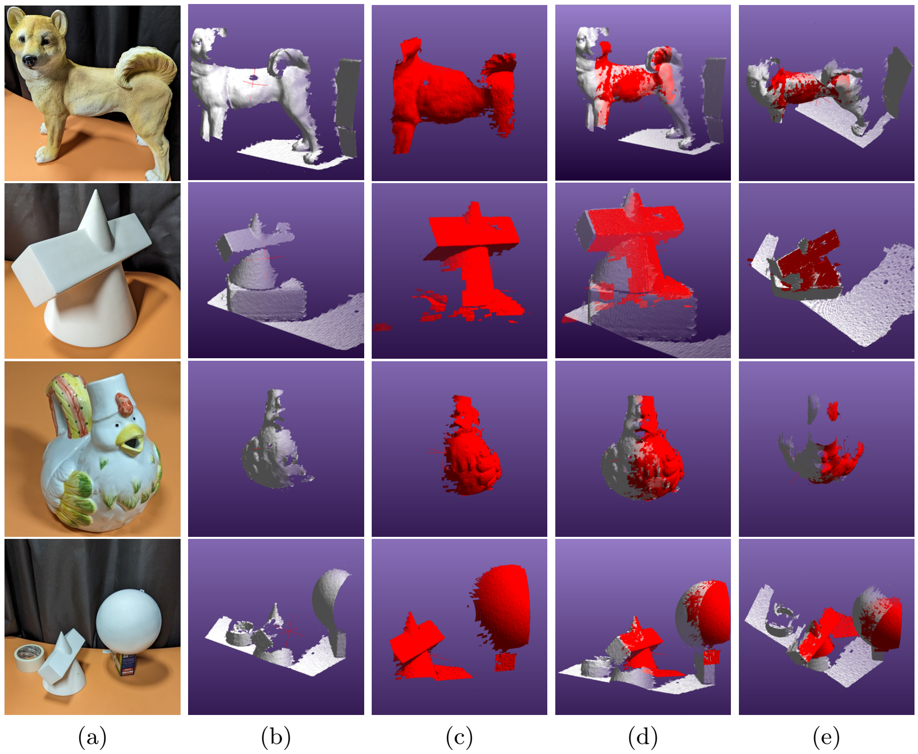


Fig. 13: Inputs and results of wide area 3D reconstruction. (a) Reference images. (b) Reconstructed shapes of objects of projector-camera system #1. (c) Reconstructed shapes of objects of projector-camera system #2. (d) Integrated shapes of system #1 and #2. (e) Integrated shapes of birds eye view. From the 4th and 5th column, it is confirmed that wider shapes are successfully reconstructed by our method. Note that shapes are integrated without using iterative closest point (ICP) [1] or similar algorithm, but automatically aligned by our auto-calibration technique using Neural-SDF.

number of cameras and projectors to measure the entire shape of an object at one time.

Acknowledgments This work was part supported by JST Startup JPMJSF23DR, ACT-X JPMJAX23C2 and JSPS/KAKENHI JP20H00611 and JP23H03439 in Japan.

References

1. Besl, P.J., McKay, N.D.: A method for registration of 3-d shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence **14**(2), 239–256 (1992). https://doi.org/10.1109/34.121791 14
2. Blais, F.: Review of 20 years of range sensor development. Journal of Electronic Imaging **13**(1), 231–243 (2004) 2
3. Chen, L., Tang, W., John, N.W., Wan, T.R., Zhang, J.J.: Slam-based dense surface reconstruction in monocular minimally invasive surgery and its application to augmented reality. Computer methods and programs in biomedicine **158**, 135–146 (2018) 3
4. DeTone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: Self-supervised interest point detection and description. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 337–33712 (2018). https://doi.org/10.1109/CVPRW.2018.00060 3
5. Furukawa, R., Mikamo, M., Sagawa, R., Kawasaki, H.: Single-shot dense active stereo with pixel-wise phase estimation based on grid-structure using cnn and correspondence estimation using gcn. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 4001–4011 (January 2022) 4
6. Furukawa, R., Oka, S., Kotachi, T., Okamoto, Y., Tanaka, S., Sagawa, R., Kawasaki, H.: Fully auto-calibrated active-stereo-based 3d endoscopic system using correspondence estimation with graph convolutional network. In: EMBC. pp. 4357–4360. IEEE (2020) 3, 5, 7, 11, 12, 13
7. Golomb, S.W., Gong, G.: Signal Design for Good Correlation: For Wireless Communication, Cryptography, and Radar. Cambridge University Press. https://doi.org/10.1017/CBO9780511546907 6
8. Gropp, A., Yariv, L., Haim, N., Atzmon, M., Lipman, Y.: Implicit geometric regularization for learning shapes. arXiv preprint arXiv:2002.10099 (2020) 9
9. Henderson, P., Ferrari, V.: Learning to generate and reconstruct 3d meshes with only 2d supervision. arXiv preprint arXiv:1807.09259 (2018) 3
10. Ikeuchi, K., Matsushita, Y., Sagawa, R., Miyazaki, D., Mukaigawa, Y., Furukawa, R., Kawasaki, H.: Active Lighting and Its Application for Computer Vision –40 Years of History of Active Lighting Techniques–. Springer (2020) 2
11. Kato, H., Ushiku, Y., Harada, T.: Neural 3d mesh renderer. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3907–3916 (2018) 3
12. Kawasaki, H., Furukawa, R., Sagawa, R., Yagi, Y.: Dynamic scene shape reconstruction using a single structured light pattern. In: CVPR. pp. 1–8. Ieee (2008) 2
13. Lamarca, J., Parashar, S., Bartoli, A., Montiel, J.: Defslam: Tracking and mapping of deforming scenes from monocular sequences. IEEE Transactions on robotics **37**(1), 291–303 (2020) 3
14. Leonard, S., Sinha, A., Reiter, A., Ishii, M., Gallia, G.L., Taylor, R.H., Hager, G.D.: Evaluation and stability analysis of video-based navigation system for functional endoscopic sinus surgery on in vivo clinical data. IEEE transactions on medical imaging **37**(10), 2185–2195 (2018) 3
15. Liao, J., Cai, L.: A calibration method for uncoupling projector and camera of a structured light system. In: 2008 IEEE/ASME International Conference on Advanced Intelligent Mechatronics. pp. 770–774. IEEE (2008) 3

16. Liu, H.T.D., Tao, M., Jacobson, A.: Paparazzi: surface editing by way of multi-view image processing. ACM Trans. Graph. **37**(6), 221–1 (2018) 3
17. Mahmoud, N., Collins, T., Hostettler, A., Soler, L., Doignon, C., Montiel, J.M.M.: Live tracking and dense reconstruction for handheld monocular endoscopy. IEEE transactions on medical imaging **38**(1), 79–89 (2018) 3
18. Matsuda, N., Cossairt, O., Gupta, M.: Mc3d: Motion contrast 3d scanning. In: 2015 IEEE International Conference on Computational Photography (ICCP). pp. 1–10 (2015). https://doi.org/10.1109/ICCPHOT.2015.7168370 3
19. Microsoft: Xbox 360 Kinect (2010), https://en.wikipedia.org/wiki/Kinect 2
20. Microsoft: Azure Kinect, (2020), https://en.wikipedia.org/wiki/Azure_Kinect 2
21. Mikamo, M., Furukawa, R., Oka, S., Kotachi, T., Okamoto, Y., Tanaka, S., Sagawa, R., Kawasaki, H.: Active stereo method for 3d endoscopes using deep-layer gcn and graph representation with proximity information. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC). pp. 7551–7555 (2021). https://doi.org/10.1109/EMBC46164.2021.9629696 3
22. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021) 3
23. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM Transactions on Graphics (ToG) **41**(4), 1–15 (2022) 9
24. Newcombe, R., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A., Kohli, P., Shotton, J., Hodges, S., Fitzgibbon, A.: Kinectfusion: Real-time dense surface mapping and tracking. In: IEEEISMAR (2011) 11, 13
25. Palazzi, A., Bergamini, L., Calderara, S., Cucchiara, R.: End-to-end 6-dof object pose estimation through differentiable rasterization. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops. pp. 0–0 (2018) 3
26. Proesmans, M., Van Gool, L.: One-shot 3d-shape and texture acquisition of facial data. In: Audio-and Video-based Biometric Person Authentication. pp. 411–418. Springer (1997) 3
27. Sarlin, P.E., DeTone, D., Malisiewicz, T., Rabinovich, A.: SuperGlue: Learning feature matching with graph neural networks. In: CVPR (2020) 3
28. Song, J., Wang, J., Zhao, L., Huang, S., Dissanayake, G.: Mis-slam: Real-time large-scale dense deformable slam system in minimal invasive surgery based on heterogeneous computing. IEEE Robotics and Automation Letters **3**(4), 4068–4075 (2018) 3
29. Ulusoy, A.O., Calakli, F., Taubin, G.: One-shot scanning using de bruijn spaced grids. In: ICCV Workshops. pp. 1786–1792. IEEE (2009) 2, 3
30. Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W.: Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. NeurIPS (2021) 3, 7, 8
31. Willomitzer, F., Häusler, G.: Single-shot 3d motion picture camera with a dense point cloud. Opt. Express **25**(19), 23451–23464 (Sep 2017). https://doi.org/10.1364/OE.25.023451, https://opg.optica.org/oe/abstract.cfm?URI=oe-25-19-23451 2
32. Yamauchi, K., Saito, H., Sato, Y.: Calibration of a structured light system by observing planar object from unknown viewpoints. In: ICPR. pp. 1–4. IEEE (2008) 3, 12, 13
33. Yamazaki, S., Mochimaru, M., Kanade, T.: Simultaneous self-calibration of a projector and a camera using structured light. In: CVPR Workshop on Projector-Camera Systems (ProCams). pp. 60–67. IEEE (2011) 11, 13

34. Zagoruyko, S., Komodakis, N.: Learning to compare image patches via convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4353–4361 (2015) 3
35. Žbontar, J., LeCun, Y.: Stereo matching by training a convolutional neural network to compare image patches. The journal of machine learning research **17**(1), 2287–2318 (2016) 3
36. Zhou, H., Jayender, J.: Emdq-slam: Real-time high-resolution reconstruction of soft tissue surface from stereo laparoscopy videos. In: MICCAI. pp. 331–340. Springer (2021) 3