Multi-scale CNN stereo and pattern removal technique
for underwater active stereo system

Kazuto Ichimaru† \hspace{1em} Ryo Furukawa‡ \hspace{1em} Hiroshi Kawasaki†
† Kyushu University, Fukuoka, Japan
‡ Hiroshima City University, Hiroshima, Japan

Abstract

Demands on capturing dynamic scenes of underwater environments are rapidly growing. Passive stereo is applicable to capture dynamic scenes, however the shape with textureless surfaces or irregular reflections cannot be recovered by the technique. In our system, we add a pattern projector to the stereo camera pair so that artificial textures are augmented on the objects. To use the system at underwater environments, several problems should be compensated, i.e., refraction, disturbance by fluctuation and bubbles. Further, since surface of the objects are interfered by the bubbles, projected patterns, etc., those noises and patterns should be removed from captured images to recover original texture. To solve these problems, we propose three approaches; a depth-dependent calibration, Convolutional Neural Network (CNN)-stereo method and CNN-based texture recovery method. A depth-dependent calibration is our analysis to find the acceptable depth range for approximation by center projection to find the certain target depth for calibration. In terms of CNN stereo, unlike common CNN-based stereo methods which do not consider strong disturbances like refraction or bubbles, we designed a novel CNN architecture for stereo matching using multi-scale information, which is intended to be robust against such disturbances. Finally, we propose a multi-scale method for bubble and a projected-pattern removal method using CNNs to recover original textures. Experimental results are shown to prove the effectiveness of our method compared with the state of the art techniques. Furthermore, reconstruction of a live swimming fish is demonstrated to confirm the feasibility of our techniques.

1. Introduction

There are strong demands on capturing dynamic scenes of underwater environments, e.g., measurement of seabeds, capturing dynamic shape deformations of swimming fish or humans, inspection of water-filled nuclear tanks by autonomous robots, etc. Passive stereo is a common solution for capturing 3D shapes because of its great advantage of simplicity; i.e., it only requires two cameras in theory. In addition, since the shapes are recovered only from a pair of stereo images, it can capture moving or deforming objects. One severe problem on passive stereo is instability, i.e., it fails to capture objects with textureless surfaces or irregular reflection. To overcome the problem, using a pattern projector to add an artificial texture onto the objects has been proposed [13]. In the system, we also take the same approach to achieve robust and dense reconstruction.

Considering underwater environments, there are additional problems for shape reconstruction by stereo, such as refraction and disturbances by fluctuation and bubbles. Further, since original textures of objects are interfered by projected patterns if active illumination is projected, they should be removed for obtaining both 3D shapes and textures. In this paper, we propose three approaches to solve aforementioned problems. For the refraction issue, a depth-dependent calibration where refractions are approximated by lens distortion of a center projection model is proposed [13]. In the paper, we analyze to find the acceptable depth range for the approximation and find the best depth for calibration. For the problems of disturbances by obstacles, we propose Convolutional Neural Network(CNN)-based stereo as a solution. Since captured images of underwater scenes are affected by mixtures of light attenuation caused by strong absorption of light intensity in water medium and strong disturbances such as bubbles, shadows of water surface or fluctuation, it is impossible to decompose them analytically. To handle such difficult problems, learning-based approaches, especially CNN techniques, are proposed.

Our shape reconstruction consists of two techniques, such as CNN-based object segmentation and CNN-based stereo matching. The CNN-based target object segmentation method efficiently segment a target object, e.g., fish in our experiment, from background, which is not only useful for reducing calculation times, but also effective to achieve robust reconstruction by narrowing the search
ranges of stereo disparities. CNN-based stereo effectively works under common variations \[31\], however, there are strong disturbances at underwater environment. In case of such strong disturbances, we propose a novel architecture of CNN, which uses multi-scale information of captured images.

Another problem for underwater environment is disturbances by bubbles, water fluctuation and other effects. Recently, convolutional neural network (CNN) based stereo matching becomes popular, which is robust to irregular distortion on image set. \citet{zbontar2016} proposed a CNN-based architecture using the WIN5RB network \[18\] which outperforms others. We also use this technique, but data collection and multi-scale extension is novel. \citet{liao2017} denoised depth images both using depth image and RGB

Experimental results are shown to prove the effectiveness of our method by comparing the results with the previous method. We also conduct demonstration to show the reconstructed sequence of a swimming fish.

2. Related works

To recover shape and texture of underwater environment, many researches have been done. Main issue for underwater environment is refraction and generally two types of solution are proposed: one is geometric approach and the other is approximation-based approach. Geometric approach is based on physical models such as refractive index, distance to refraction interface, and normal of the interface. \citet{agrawal2017} introduced polynomial formulation for the model \[1\]. \citet{sedlazeck2011} proposed structure from motion for underwater environment \[11\]. \citet{kawahara2014} proposed pixel-wise varifocal camera model \[12\]. In this model, appropriate focal lengths are assigned to each pixel. Those techniques can calculate genuine light rays if parameters are correctly estimated and interface is completely planar, however, they are usually impractical. On the other hand, approximation approach converts captured images into central projection images by lens distortion and focal length adjustment \[7\]. They assumed focal point moved backward to adjust light paths as linear as possible, then remaining error was treated as lens distortion. \citet{kawasaki2013} also proposed a simple method to approximate the refraction by radial distortion \[13\]. Since the parameter cannot be fixed for all the depth range, they proposed a depth dependent technique. It works well in most cases, however in specific case it fails because refractive distortion depends on depth and effective range of depth is not thoroughly analyzed yet.

Another problem for underwater environment is disturbances by bubbles, water fluctuation and other effects. Recently, convolutional neural network (CNN) based stereo matching becomes popular, which is robust to irregular distortion on image set. \citet{zbontar2016} proposed a CNN-based architecture using the WIN5RB network \[18\] which outperforms others. We also use this technique, but data collection and multi-scale extension is novel. \citet{liao2017} denoised depth images both using depth image and RGB

For the texture recovery, we also propose a CNN-based method for projected-pattern removal and bubble cancellation. Main contributions of the proposed technique are as follows:

1. A practical technique is proposed to achieve dense and robust shape reconstruction based on passive stereo using active pattern projection.
2. A valid depth range for depth-dependent approximation by radial distortion is analysed.
3. A target-region detection method by CNN for robust stereo matching is proposed.
4. A multi-scale CNN-based stereo technique specialized for underwater environment is proposed.
5. A multi-scale CNN-based bubble and projected pattern removal method specialized for underwater environment is proposed.

Experimental results are shown to prove the effectiveness of our method by comparing the results with the previous method. We also conduct demonstration to show the reconstructed sequence of a swimming fish.
image. Zhang et al. [32] denoised images with CNNs with different noise levels taken into account. Choi et al. [5] proposed denoising with multi-scales with light-weight computation. Nakamura et al. [23] removed texts in natural scene images using multi-layers of convolutions and deconvolutions.

Image segmentation is also important for our system, since usually only the regions of the target object are enough for 3D shape reconstruction. Badrinarayanan et al. proposed a network architecture for semantic segmentation called SegNet [2]. Ronneberger et al. also proposed a network architecture called U-Net which is useful for biomedical image segmentation [24]. Since captured images do not look similar to scenery image, but rather close to biomedical image, we use U-Net for our segmentation.

3. System and algorithm overview

3.1. System Configuration

Our system consists of stereo camera pair and one laser projector as shown in Fig. 1. We prepare two systems for our experiments. One is for evaluation purpose where two cameras and a projector are set outside a water tank. The other is a practical system where devices are installed into a specially built waterproof housing in order to make distance between interface glass and camera lens to be relatively short. For the both systems, the optical axes of the cameras are set orthogonal to glass surface so that error by refraction approximation is minimized. The two cameras are synchronized by GPIO cable to capture dynamic scenes. In terms of the pattern projector to add textures onto the objects, no synchronization is required since the pattern is static. In our implementation, we use a laser projector where diffractive optic element (DOE) is used to configure wave pattern proposed in [25] without losing light power.

3.2. Algorithm

The algorithm of our underwater shape reconstruction will be explained by using Fig. 2. First, the camera pair is calibrated. The refractions in the captured images are modeled and canceled by center projection approximation in our technique using depth-dependent intrinsic and extrinsic parameters which are acquired in advance. In the measurement process, the targets are captured with stereo cameras. Pattern illumination is projected onto the scene for adding features on it. From captured images, target regions are detected by a CNN-based segmentation technique, where only fish regions are extracted. Then, a stereo-matching method is applied to the target regions. In our technique, a CNN-based stereo is applied to increase stability under the condition of dimmed patterns, disturbances by bubbles, and flickering shadows. Then, 3D points are reconstructed from the disparity maps estimated by the stereo algorithm. Outliers are removed from the point cloud and meshes are recovered by Poisson equation method [14]. Since textures are degraded by bubbles and projected patterns, they are efficiently recovered by CNN-based bubble canceling and pattern removal techniques. Using the recovered 3D shapes and textures, we can render the dynamic and textured 3D scene.

3.3. Depth-dependent calibration

Because of refractions, captured images of underwater scene are severely distorted. In this paper, we undistort captured images by a lens distortion model [13]. The technique is only an approximation, because refraction effect is not strictly represented as the lens-distortion model, but it can be used for stereo matching for limited working dis-
For the actual process, a calibration tool, i.e., planar board with checker pattern, is submerged to a water tank to retrieve intrinsic and extrinsic (camera-to-camera transformation) parameters, and thus, it is preferable if the best depth for approximation is known in advance. In the paper, we simulate error using actual parameter of our system as shown in Fig. 3, showing that a maximum error is below 0.8% if depth range is less than 1 m. Thus, we set all the devices as close as possible to the water interface so that the error becomes small enough to be ignored.

4. CNN based stereo technique with pattern projection

In the technique, we first apply CNN-based target region extraction technique (Sec. 4.1) to increase robustness as well as decrease calculation time. Then, multi-scale CNN stereo (Sec. 4.3) is applied to reconstruct 3D points.

4.1. CNN-based target-region extraction

For many applications, reconstruction targets are recognizable, such as swimming fishes in the water. In general, the wider the range of disparities considered in stereo-matching processes, the more ambiguities exist, leading to wrong correspondences. Thus, by extracting the target regions from the input images and reducing possibilities of matching within the detected target regions, 3D reconstruction process becomes more robust.

To this purpose, we implemented an U-Net [24], an FCN with multi-scale feature extraction, and trained it. We made training dataset from underwater image sequence contains live fish (since one of our applications is live fish measurement) where scenes are illumination by the pattern projector. Since both the target and background regions are projected with the same pattern, segmentation between those regions was difficult. From image sequences, 100 images were sampled and the target regions were masked with manual operations. These training images were augmented by scalings, rotations, and translations. As a result, we provide 980 pairs of source images and target-region masks for training U-Net. We used softmax entropy for loss function. The trained U-Net was tested for large number of images, we obtained qualitatively successful results in most examples (Fig. 3).

In the evaluation process, we have found that the numbers of resolution levels of the U-Net architecture is important. By using only two or three levels of resolutions, we could not get sufficient results. We finally reached the conclusion that the U-Net with five levels of resolutions works effectively with our dataset by qualitative evaluation increasing number of levels. Regarding the number of training data pairs, 300 augmented image pairs from around 30 annotated data did not work sufficiently for a living fish, but at least 100 pairs were required.

4.2. CNN stereo matching by transfer learning

In general, normal stereo-matching methods such as SGBM are not robust against strong noises since they do not classify pixels into right intensity and wrong intensity [9]. Because CNN-based stereo proposed in [31] learns from real images, it is possible to cope with the noises. In the technique, small image patches from stereo image pairs are processed by CNNs and their feature vectors are calculated. Similarity measures of the feature vectors are used to find the best-matching disparities for every patches of the input images.

In the method, we propose an effective training method for CNN-based stereo specialized for bubble-disturbed images by applying a transfer-learning technique. First, we made a training dataset disturbed by bubbles from Middlebury 2005 and 2006 dataset. Middlebury dataset contains 1,890 images in total, and we used 540 images of them. To create images with bubbles, we set a display monitor behind a water tank and put a bubble generator inside a tank (Fig. 5 (left)). The Middlebury images were presented on the monitor and captured by the camera in front of the water tank. The captured images were warped both by the perspective projection and the refraction by the air-water interfaces. To compensate for this, gray code was presented on the display screen and captured by the camera. Then, lens distortion parameters are estimated, which approximate the refraction, by using the gray code. The captured images were undistorted by the lens distortion parameters and rectified by homography transformation. Examples of a source image and their bubble-disturbed images are shown in Fig. 5 (right).

Since Middlebury dataset is annotated with ground-truth disparities, we can get positive and negative pairs of image patches for stereo-matching training data. The positive pairs of patches are sampled from stereo images with corresponding positions, whereas the negative pairs are sampled randomly. Using these matching pair datasets, we additionally trained the CNN-based similarity measure pipeline with the captured dataset.
input patches were explicitly extracted from the same epipolar rotations, scalings, and brightness changes. Note that similarly with section 4.2 with data augmentation of randomly, the training dataset created from modified (i.e., with bubble) Middlebury dataset similarly with section 4.2 with data augmentation of random rotations, scalings, and brightness changes. Note that input patches were explicitly extracted from the same epipolar lines of input images in training phase, but whole image can be inputted in estimation phase.

5. Texture recovery from noise, bubble and projected pattern

For real situations, the captured images are often severely degraded by underwater environments, such as bubble and other noises, as well as projected pattern on the object surface. In order to remove such undesirable effects, we propose a CNN-based texture recovery technique. In our technique, we focus on two major problems, such as bubbles and projected patterns. Although those two phenomena are totally different and have different optical attributes, it is common in the sense that appearances for both effects have a wide variation in scale. Note that such wide variation depends on the distance between a target object, bubble and a projector. Such a large variation of scale makes it difficult for removal by simple noise removal method.

Since multi-scale CNN is suitable to learn such a variation, we also use a multi-scale CNN for our bubble and pattern removal purpose. The network for such obstacle removal is shown in Fig. 6(Right). In the figure, it is shown that an original image is converted to three different resolutions and trained by independent CNN. Each output is up-sampled and concatenated to higher resolution. This network is advantageous because it can handle a large structure of projected pattern, as well as it can be trained in a relatively short time. We prepared two datasets to train the network for bubble removal and pattern removal. For training bubble removal network, we also used Middlebury dataset containing bubbles mentioned in Sec. 4.3. For training pattern removal network, we captured several real targets with/without pattern projection to create training data. However, the number of data is not sufficient to train the network, we synthesize training data by using CG. We use Middlebury dataset and reconstruct 3D shape with texture map, and then, use virtual pattern projector to add pattern onto the object surface. Then, images were translated, rotated, and scaled randomly for data augmentation. The pattern removal ability of this network is shown in the experiment.

6. Experiments

To evaluate proposed method, we conducted 4 experiments. In Sec. 6.1 we describe how our method is accurate and dense under depth-dependent calibration. In Sec. 6.2, it is examined that how our multi-scale CNN stereo is robust against underwater disturbances. In Sec. 6.3, qualitative evaluation results of texture recovery are shown. Finally in Sec. 6.4, we captured and reconstructed real swimming fish to confirm the feasibility of our method.
6.2. Evaluation of various CNN stereo techniques

Next, we tested CNN-based stereo for underwater scene with bubbles. For evaluation purpose, we prepared four implementations, such as CNN-based stereo of [31], multi-scale CNN stereo with linear combination (ms-cnn-lin), multi-scale CNN stereo with FCN (ms-cnn-fcn), and transfer learned ms-cnn-lin with bubble erased images (ms-cnn-lin(trans)). The target objects were placed at a distance of 50, 60, 70cm and the depth-dependent calibration was applied as same as the previous experiment. We intentionally made bubbles to interfere image capturing process. We reproduced four bubble environments, i.e., far little bubble, far much bubble, near little bubble, and near much bubble. In addition, no bubble scenes as reference were prepared. We captured three pairs of images for each target with five environments. In total, 90 images were captured. Then, we removed bubbles on the images with multi-scale bubble removal architecture, and reconstructed all the scenes and targets. We calculated average RMSE from the GT shape of each target. The results are shown in Fig. 10. From the graph, we can confirm that the accuracy of proposed CNN architecture is better than previous method, supporting the effectiveness of our method. Fig. 11 shows examples of the reconstructed disparity maps (masked with segmentation results) for each technique confirming that shapes are recovered by our technique even if captured images are severely degraded by bubbles.

6.3. Experiments of texture acquisition

We also tested the bubble-removal and the pattern-removal techniques. The results are shown in Fig. 12. It is shown that bubbles in the source images were successfully removed as shown in the top row of the figure. In the bottom row, we can also confirm that projected patterns are robustly removed by multi-scale CNN technique.

6.4. Demonstration with a live fish

Finally, we captured a live swimming fish (filefish) at an aquarium. We used a special experimental system with an
aluminum housing. Cameras are same as the above experiment, but a projector is substituted by laser pattern projec-
tor. We captured and reconstructed 360 frames. Five frames from the results are shown in Fig. 8 for example. As shown in the figure, we can confirm that the target object is mostly successfully segmented by our CNN-based object segmentation method. In addition, dense shapes of the swimming fish are successfully reconstructed, which proves the effectiveness and practicality of our method. Texture are also partially recovered with our method.

7. Conclusion

The paper presents a practical underwater dense shape reconstruction technique as well as texture refinement method using stereo cameras with a static-pattern projector. Since underwater environments have severe conditions, such as refraction, light attenuation and disturbances by bubbles, we propose a CNN-based solutions, such as a target-object segmentation, robust stereo matching with a multi-scale CNN and CNN based texture-recovery method. By comparing 3D shape reconstruction with various methods, since other methods are severely affected by bubbles and other degradation of underwater environment, our method achieved best among them. Further, bubbles and projected patterns on the objects are successfully removed by our method. We also conducted experiments to show that our approximation of refraction by radial distortion is feasible. Our future plan is to apply the technique to a swimming human for sports analysis.

Acknowledgment

This work was part supported by grant JSPS/KAKENHI 16H02849, 16KK0151, 18H04119, 18K19824 in Japan, and MSRA CORE14.
Figure 11. Difference of disparity maps between stereo methods in bubble scene. Bubble is so severe and almost any method can produce quite poor results, whereas our method produced much better results.

Figure 12. Result of texture acquisition experiment. Left pane is results of bubble removal and Right pane is results of pattern removal. 
Left: Input images. Middle-left: Close-up view of input. Middle-right: Output images. Right: Close-up view of output. (a, b): Middlebury dataset. (c): Fish model. (d, e): Pattern removal dataset we created. (f): Live fish.
References

[1] A. Agrawal, S. Ramalingam, Y. Taguchi, and V. Chari. A theory of multi-layer flat refractive geometry. In CVPR, 2012.

[2] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(2):2481–2495, January 2017.

[3] Z. Cai, Q. Fan, R. S. Feris, and N. Vasconcelos. A unified multi-scale deep convolutional neural network for fast object detection. In ECCV, 2016.

[4] J. Chen and C. Yuan. Convolutional neural network using multi scale information for stereo matching cost computation. In Image Processing (ICIP). IEEE, 2016.

[5] S. Choi, J. Isidoro, P. Getreuer, and P. Milanfar. Fast, trainable, multiscale denoising, 2018.

[6] Q. Fan, J. Yang, G. Hua, B. Chen, and D. P. Wipf. A generic deep architecture for single image reflection removal and image smoothing. In IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017, pages 3258–3267. IEEE Computer Society, 2017.

[7] R. Ferreira, J. P. Costeira, and J. A. Santos. Stereo reconstruction of a submerged scene. In Pattern Recognition and Image Analysis, pages 102–109. Springer, 2005.

[8] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 12 2016.

[9] H. Hirschmüller. Stereo processing by semiglobal matching and mutual information. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2):328–341, February 2008.

[10] S. Iizuka, E. Simo-Serra, and H. Ishikawa. Globally and locally consistent image completion. ACM Trans. Graph., 36(4):107:1–107:14, July 2017.

[11] A. Jordt-Sedlazeck and R. Koch. Refractive structure-from-motion on underwater images. In Computer Vision (ICCV), 2013 IEEE International Conference on, pages 57–64. IEEE, 2013.

[12] R. Kawahara, S. Nobuhara, and T. Matsuyama. A pixel-wise varifocal camera model for efficient forward projection and linear extrinsic calibration of underwater cameras with flat housings. In Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on, pages 819–824. IEEE, 2013.

[13] H. Kawasaki, H. Nakai, H. Baba, R. Sagawa, and R. Furukawa. Calibration technique for underwater active oneshot scanning system with static pattern projector and multiple cameras. In Winter Conference on Applications of Computer Vision. IEEE, 2017.

[14] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In Proceedings of the fourth Eurographics symposium on Geometry processing, SGP ’06, pages 61–70, 2006.

[15] K. Konolige. Projected texture stereo. In International Conference on Robotics and Automation, pages 148–155. IEEE, 2010.

[16] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang. Reside: A benchmark for single image dehazing. arXiv preprint arXiv:1712.04143, 2017.

[17] X. Liao and X. Zhang. Multi-scale mutual feature convolutional neural network for depth image denoise and enhancement. IEEE, 2017.

[18] P. Liu and R. Fang. Wide inference network for image denoising. CoRR, abs/1707.05414, 2017.

[19] H. Lu, H. Xu, L. Zhang, and Y. Zhao. Cascaded multi-scale and multi-dimension convolutional neural network for stereo matching, 2018.

[20] W. Luo, A. G. Schwing, and R. Urtasun. Efficient deep learning for stereo matching. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 5695–5703, June 2016.

[21] N. Mayer, E. Ilg, P. Häusser, P. Fischer, D. Cremers, A. Dosovitskiy, and T. Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016, pages 4040–4048. IEEE Computer Society, 2016.

[22] S. Nah, T. H. Kim, and K. M. Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[23] T. Nakamura, A. Zhu, K. Yanai, and S. Uchida. Scene text eraser. In The 14th International Conference on Document Analysis and Recognition, pages 2–45, 2017.

[24] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241. Springer, 2015.

[25] R. Sagawa, K. Sakashita, N. Kasuya, H. Kawasaki, R. Furukawa, and Y. Yagi. Grid-based active stereo with single-colored wave pattern for dense one-shot 3D scan. In 3DIM-PVT, pages 363–370, 2012.

[26] A. Shaked and L. Wolf. Improved stereo matching with constant highway networks and reflective confidence learning. In 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, pages 6901–6910. IEEE Computer Society, 2017.

[27] A. Tonioni, M. Poggi, S. Mattoccia, and L. D. Stefano. Unsupervised adaptation for deep stereo. In 2017 IEEE International Conference on Computer Vision (ICCV), volume 00, pages 1614–1622, Oct. 2018.

[28] S. Tulyakov, A. Ivanov, and F. Fleuret. Weakly supervised learning of deep metrics for stereo reconstruction. In 2017 IEEE International Conference on Computer Vision (ICCV), pages 1348–1357, Oct 2017.

[29] P. Yadati and A. M. Namboodiri. Multiscale two-view stereo using convolutional neural networks for unrectified images. In 15th IAPR International Conference on Machine Vision Applications (MVA), pages 320–323. IEEE, 2017.

[30] C. You, Q. Yang, H. Shan, L. Gjesteby, G. Li, S. Ju, Z. Zhang, Z. Zhao, Y. Zhang, W. Cong, and G. Wang. Structure-sensitive multi-scale deep neural network for low-dose ct denoising, 2018.
[31] J. Žbontar and Y. LeCun. Stereo matching by training a convolutional neural network to compare image patches. *Journal of Machine Learning Research*, 17:2287–2318, January 2016.

[32] K. Zhang, W. Zuo, and L. Zhang. Fdnet: Toward a fast and flexible solution for cnn based image denoising, 2017.

[33] C. Zhou, H. Zhang, X. Shen, and J. Jia. Unsupervised learning of stereo matching. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 1576–1584, Oct 2017.