

Deep compressive sensing with coded image sensor

Michitaka Yoshida^{1,2}, Daisuke Hayashi², Lioe De Xing², Keita Yasutomi²,
Shoji Kawahito², Keiichiro Kagawa², Hajime Nagahara³

1 Japan Society for the Promotion of Science
3-5-1 Johoku, Hamamatsu, 432-8011 Japan

2 Shizuoka University

3-5-1 Johoku, Hamamatsu, 432-8011 Japan

3 Osaka University

2-8, Yamadaoka, Suita, Osaka 565-0871 Japan

E-mail: yoshida.michitaka@shizuoka.ac.jp

Abstract In this paper, we introduce a method of compressed sensing using coded CMOS sensors and the concept of deep sensing. By shifting the exposure timing for each pixel, temporal information can be encoded into a single image, which can then be applied to various tasks by recovering the temporal information through post-processing. Deep sensing is the concept of using an end-to-end neural network to represent sensing and reconstruction in compressed sensing. By jointly optimizing the sensing and reconstruction processes, deep sensing enhances reconstruction quality by enabling more efficient sensing compared to random sampling. We introduce the effectiveness of this approach through applications such as video compressed sensing, human action recognition from coded exposure images, compressed light field observation, and compressed transient image observation with depth map estimation.

Keywords: Compressive sensing, Deep optics, Coded exposure image sensor

1. High-speed imaging with compressive sensing using coded exposure CMOS image sensor

Compressed sensing methods using temporally coded exposure have been proposed to improve the temporal resolution of image sensors [1-3]. Traditionally, high-frame-rate videos are captured by high-speed cameras, but these special sensors are expensive, have low spatial resolution, and have poor sensitivity. To solve this problem, a method has been proposed to encode temporal information in a single image by shifting the exposure timing for each pixel and to reconstruct a video from the captured image in post-processing.

2. Human action recognition using coded exposure images

A method of reconstruction-free action recognition from a single coded exposure image has been proposed [4,5]. For a camera, there is a trade-off between spatial resolution and frame rate. A feasible approach to overcome this trade-off is compressive video sensing. Compressive video sensing uses random coded exposure and reconstructs higher than read out of sensor frame rate video from a single coded image. It is possible to recognize an action in a scene from a single coded image because the image contains multiple temporal information for reconstructing a video. Unlike ordinary images, encoded exposure images are not suitable for deep learning using convolution because neighboring pixels have different information. A convolution layer with a shift-displacement kernel was used to solve this problem.

3. Dynamic light field imaging with coded aperture and coded exposure

Mizuno et al. propose a method for compressively acquiring a dynamic light field (a 5-D volume) through a single-shot coded image (a 2-D measurement) [6,7]. They designed an imaging model that synchronously applies aperture coding and pixel-wise exposure coding within a single exposure time. This coding scheme enables us to effectively embed the original

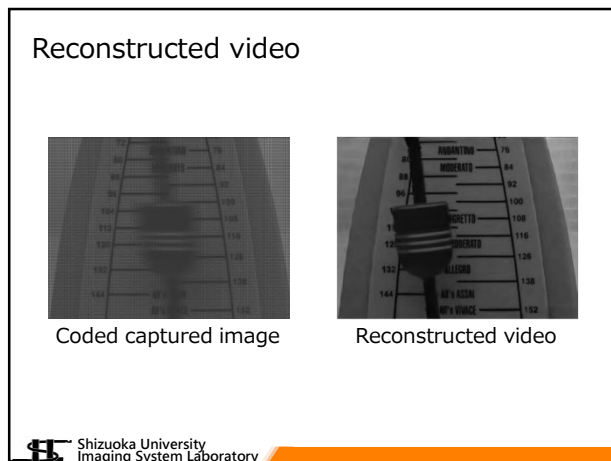
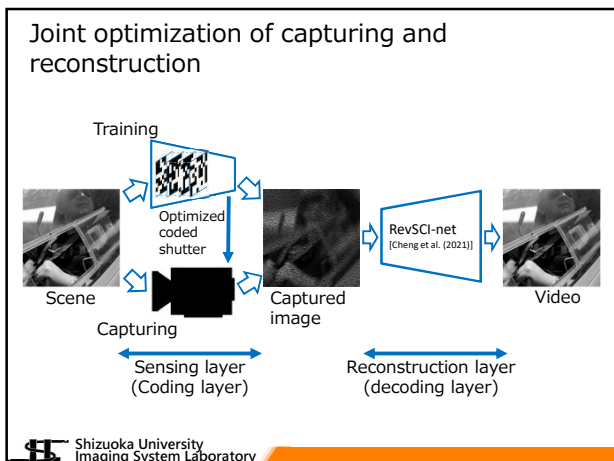
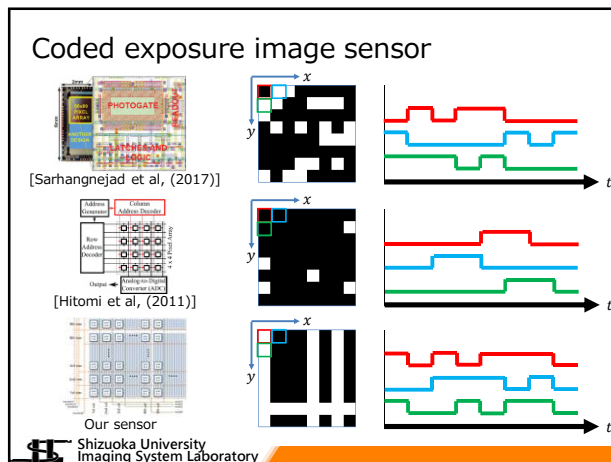
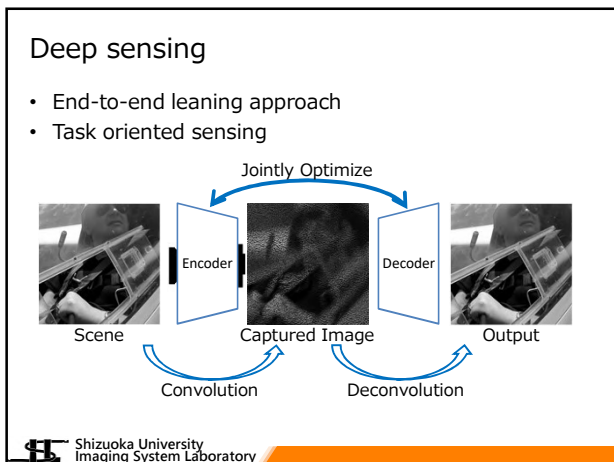
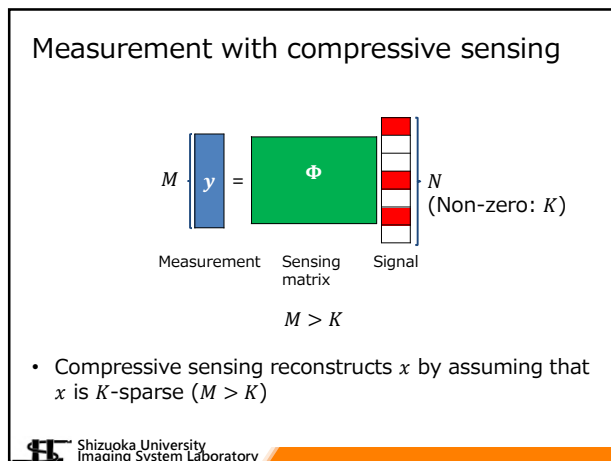
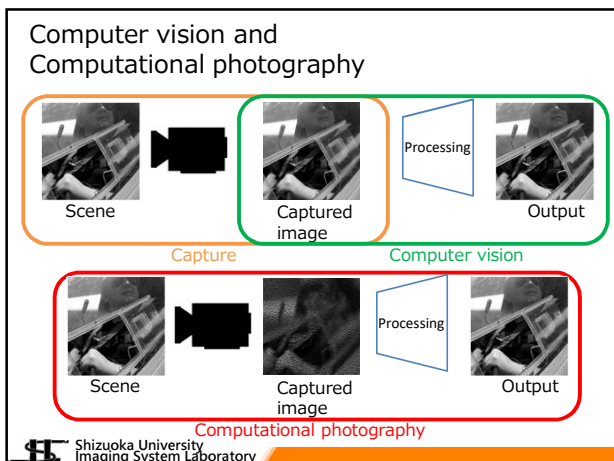
information into a single observed image. The observed image is then fed to a convolutional neural network (CNN) for light-field reconstruction, which is jointly trained with the camera-side coding patterns. They also developed a hardware prototype to capture a real 3-D scene moving over time. They succeeded in acquiring a dynamic light field with 5x5 viewpoints over 4 temporal sub-frames (100 views in total) from a single observed image. Repeating capture and reconstruction processes over time, a dynamic light field can be acquired at 4x the frame rate of the camera.

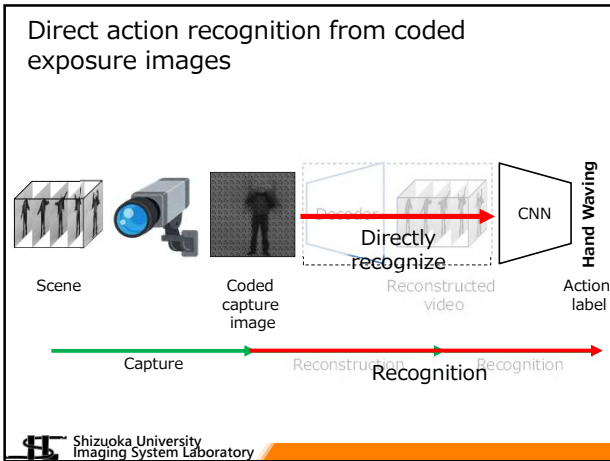
4. Compressive transient imaging and depth estimation with optimal coded shutter

Transient imaging, which uses a multi-tap coded exposure CMOS image sensor to observe the propagation of light through a time-coded exposure, has been proposed [9,10]. Conventional transient imaging requires a special sensor and has problems with resolution, SNR, and measurement time. To solve this problem, a multi-tap encoded shutter CMOS sensor [8] was used to acquire transient images by compressed sensing. A multi-tap CMOS sensor is a sensor that has multiple circuits (taps) to store the charge generated by photodiodes, and achieves high temporal resolution by switching the taps to be transferred at high speed. However, a uniform shutter cannot observe enough temporal information, so enough temporal information is convolved into each tap by temporal encoding and observed. By reconstructing the transient image from observations, we reduce the MPI effect in ToF imaging. This method can be applied to dynamic scenes because a coded shutter sensor can observe the scene response in a single shot. Furthermore, we improve depth estimation accuracy by optimizing the reconstruction network with a coded shutter that compresses the scene response. Furthermore, the coded shutter that compresses the scene response is optimized simultaneously with the reconstruction network to improve reconstruction quality and depth estimation accuracy.

References

- [1]Yoshida Michitaka, Torii Akihiko, Okutomi Masatoshi, Endo Kenta, Sugiyama Yukinobu, Taniguchi Rin-ichiro, Nagahara Hajime, “Joint optimization for compressive video sensing and reconstruction under hardware constraints”, Proceedings of the European Conference on Computer Vision (ECCV2018), Munich, Germany, pp.649-663, 2018.09
- [2]Michitaka Yoshida, Toshiki Sonoda, Hajime Nagahara, Kenta Endo, Yukinobu Sugiyama, Rin-ichiro Taniguchi, “High-Speed Imaging Using CMOS Image Sensor With Quasi Pixel-Wise Exposure”, IEEE Transactions on Computational Imaging, Vol.6, pp.463-476, 2019.
- [3]Yoshida Michitaka, Akihiko Torii, Masatoshi Okutomi, Rin-ichiro Taniguchi, Hajime Nagahara, and Yasushi Yagi, “Deep Sensing for Compressive Video Acquisition”, Sensors 2023, 23, no. 17: 7535..
- [4]Tadashi Okawara, Michitaka Yoshida, Hajime Nagahara, Yasushi Yagi, “Action Recognition from a Single Coded Image”, Proceedings of the International Conference on Computational Photography (ICCP2020), Saint Louis, U.S.A, 2020.04.
- [5]Sudhakar Kumawat, Tadashi Okawara, Michitaka Yoshida, Hajime Nagahara, Yasushi Yagi, “Action Recognition From a Single Coded Image”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 4, pp. 4109-4121, 1 April 2023.
- [6]Ryoya Mizuno, Keita Takahashi, Michitaka Yoshida, Chihiro Tsutake, Toshiaki Fujii, Hajime Nagahara, “Acquiring a Dynamic Light Field Through a Single-Shot Coded Image”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, U.S.A, 2022.06.
- [7]Ryoya Mizuno, Keita Takahashi, Michitaka Yoshida, Chihiro Tsutake, Toshiaki Fujii, and Hajime Nagahara, “Compressive Acquisition of Light Field Video Using Aperture-Exposure-Coded Camera”, ITE Transactions on Media Technology and Applications 2024, Vol.12, Issue 1, pp.22-35.
- [8]Futa Mochizuki, Keiichiro Kagawa, Shin-ichiro Okihara, Min-Woong Seo, Bo Zhang, Taishi Takasawa, Keita Yasutomi, and Shoji Kawahito, "Single-event transient imaging with an ultra-high-speed temporally compressive multi-aperture CMOS image sensor." Optics express 24.4 (2016): 4155-4176..
- [9]Michitaka Yoshida, Daisuke Hayashi, Lioe De Xing, Keita Yasutomi, Shoji Kawahito, Keiichiro Kagawa, Hajime Nagahara, “Single-shot efficient transient imaging with optimal coded shutter based on time-compressive CMOS image sensor”, Computational Optical Sensing and Imaging (COSI), Toulouse, France, 2024.07.
- [10] Michitaka Yoshida, Daisuke Hayashi, Lioe De Xing, Keita Yasutomi, Shoji Kawahito, Keiichiro Kagawa, Hajime Nagahara, “Single-shot efficient depth imaging based on time-compressive CMOS image sensor”, Proceedings of the International Conference on Computational Photography (ICCP2024), Lausanne, Switzerland, 2024.07

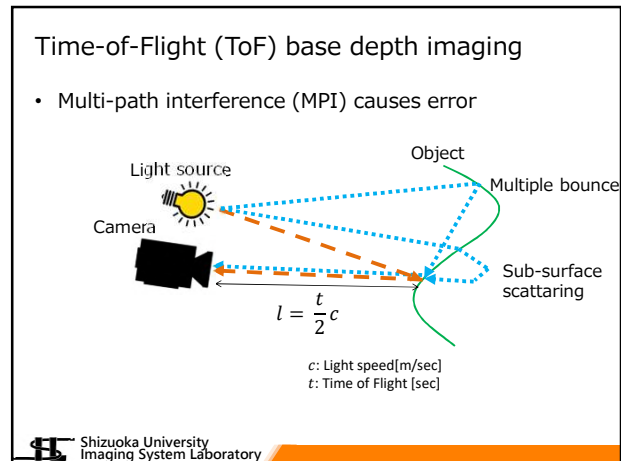
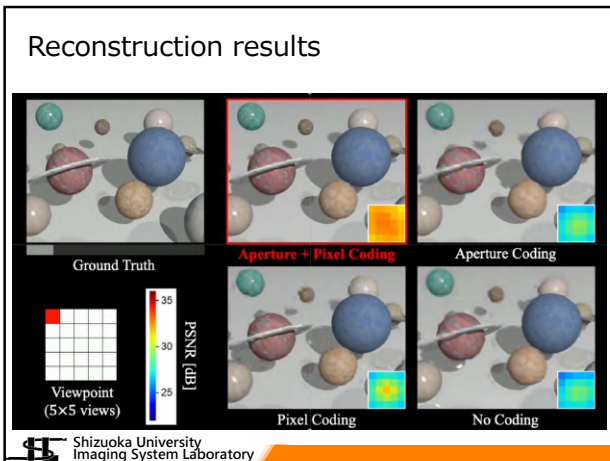
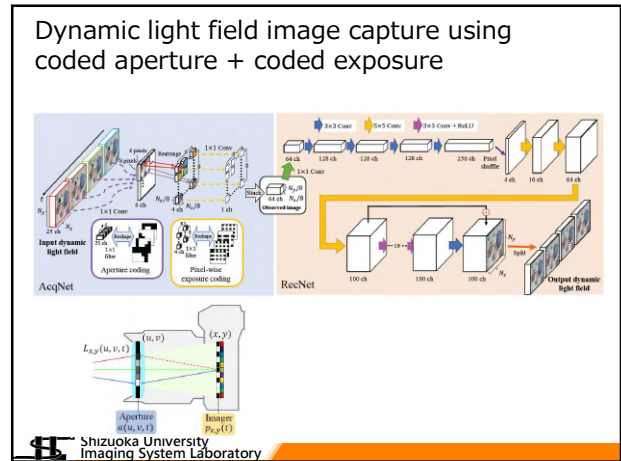
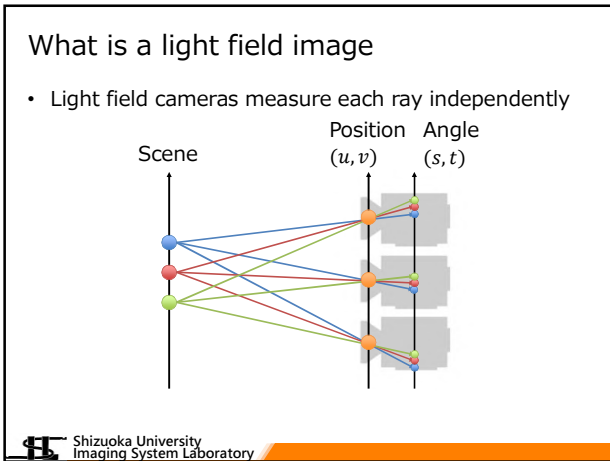




Recognition results

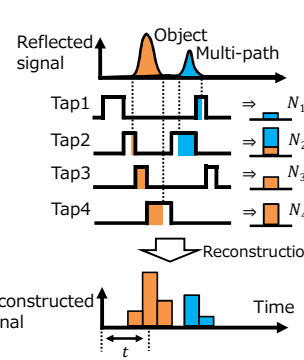
		Accuracy [%]			
Input	Model	Top1	Top3	Top5	
Video (upper bound)	C3D	39.31	61.97	70.05	
Single Image	Code exposure (Proposed)	29.37	47.39	56.33	
	Long exposure	C2D	10.82	22.83	30.20
	Short exposure	C2D	10.32	21.85	28.58

Shizuoka University Imaging System Laboratory



Compressive ToF

- Compresses signals with temporal coded shutter
 - With iToF based sensor
- Reconstruct signal from coded capture image
- Estimate depth from reconstructed signal

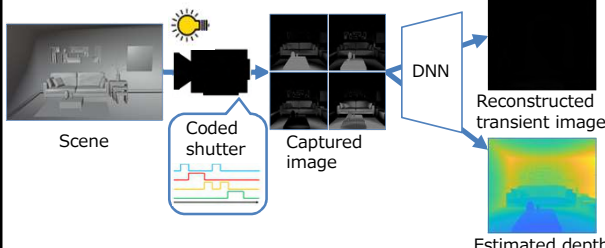


- ✓ High image resolution
- ✓ Robust to MPI

Shizuoka University Imaging System Laboratory

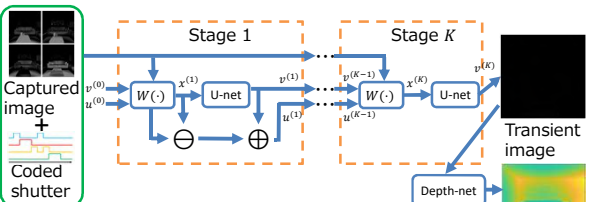
ToF imaging based on compressive sensing

- Compresses signals in the charge domain
 - by switching the taps with coded shutter
- Reconstruct the signal and estimate the depth



Shizuoka University Imaging System Laboratory

Reconstruction with ADMM-net



- ADMM : Alternate iterative optimization
 - $W(\cdot)$: Initial reconstruction with coded shutter
 - U-net : Denoising
- Loss function
 - $Loss_l = RMSE(v_K, v_{GT}) + \frac{1}{2}RMSE(v_{K-1}, v_{GT}) + \frac{1}{2}RMSE(v_{K-2}, v_{GT})$

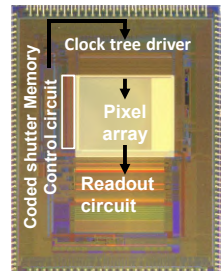
Shizuoka University Imaging System Laboratory

Simulation results

	GT	TVAL3 +Random	ADMM-net +Random	ADMM-net +Optimized
Image PSNR		25.31	30.01	30.34
Depth MSE		0.0143	0.00917	0.00356

Shizuoka University Imaging System Laboratory

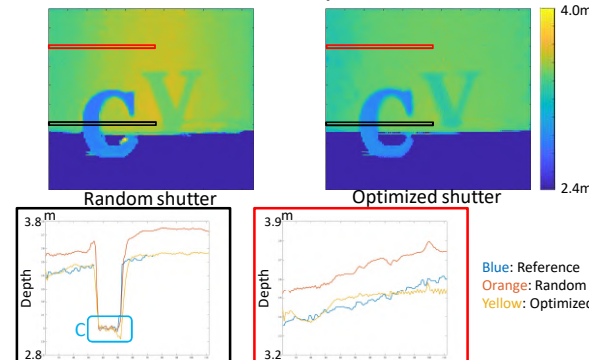
Specifications of prototype sensor



Technology	0.11 μm CIS
Burst frame rate	607 Mfps
Read frame rate	< 21 fps
Number of effective pixels (sub-pixels)	212 (H) \times 188 (V)
Pixel pitch (sub-pixel)	11.2 μm (H) \times 11.2 μm (V)
Total number of taps	16 (4 \times 4)
Chip size	7.0 mm (H) \times 9.3 mm (V)
Coding bit length	8-256

Shizuoka University Imaging System Laboratory

Estimated real scene depth



Blue: Reference
Orange: Random
Yellow: Optimized

Shizuoka University Imaging System Laboratory